

---

**PEPLOT\*, a protein secondary structure analysis program for the UWGCG sequence analysis software package**

---

Michael Gribskov<sup>1</sup>, Richard R. Burgess<sup>2</sup> and John Devereux<sup>3</sup>

---

<sup>1</sup>Molecular Biology Institute, Department of Chemistry, University of California, Los Angeles, CA,

<sup>2</sup>McArdle Laboratory for Cancer Research, University of Wisconsin, Madison, WI 53706 and

<sup>3</sup>University of Wisconsin Genetics Computer Group, UW Biotechnology Center, 1710 University Avenue, Madison, WI 53705, USA

---

Received 17 July 1985

---

**ABSTRACT**

We describe a program for the analysis of protein secondary structure that operates with the Sequence Analysis Software Package of the University of Wisconsin Genetics Computer Group (UWGCG). The program produces both graphic and printed output. Structure prediction using the Chou and Fasman and Robson et al methods, and hydropathy analysis by the method of Kyte and Doolittle are included along with a simplified method of hydrophobic moment analysis. The power of the program is the coordinated presentation of many different kinds of structural information on the same plot.

**INTRODUCTION**

We have developed a program for protein secondary structure analysis that combines several of the most frequently used techniques. Information necessary to predict protein secondary structures by the Chou and Fasman (1) method is shown graphically including plots of the alpha and beta structural potentials, reverse turn potential, and probability of alpha and beta structure ends. Structure predictions using the method of Robson et al (2) are available in printed form for several ranges of decision constants.

The hydropathy profile (3) is frequently used to find intra-membrane regions of proteins. It is also commonly used to try to predict the positions of antigenic sites on proteins (4).

A more recently developed technique for protein secondary structure prediction is the hydrophobic moment plot (5). The hydrophobic moment is a measure of the "amphiphilicity" of a sequence that is helpful for identifying structures located at the interface between hydrophobic and hydrophilic regions of a protein. These plots are usually presented as contour plots with all possible inter-residue angles on one axis and the sequence position on the

\* PEPLOT is distributed as a part of the UWGCG Sequence Analysis Software Package described in reference 6. The package is available for \$2,400 (non-profit) or \$4,800 (commercial) from John Devereux, UW Biotechnology Center, 1710 University Avenue, Madison, Wisconsin 53705, USA, (608) 263-8970.

other. We have chosen to focus on the maximum hydrophobic moment for inter-residue angles characteristic of alpha and beta structures, allowing the results to be plotted as a pair of continuous curves.

### MATERIALS AND METHODS

PEPLOT is designed to operate as part of the UWCGG Sequence Analysis Software Package (6). This package is designed for VAX computers running version 3 or 4 of the VMS operating system. The program was written in FORTRAN 77 using the UWCGG procedure library--a library of subroutines that simplifies sequence manipulations, input and output operations, and graphics programming. The UWCGG package, and PEPLOT, currently support Hewlett Packard 7221 and 7475 plotters and DEC VT240 and 241 graphic terminals.

Conformational parameters for the Chou and Fasman structural prediction were obtained from their review article (1). This article should be consulted for specific information on the Chou and Fasman prediction method.

The method of Robson et al is the version presented by Garnier et al (2). Predictions with several possible combinations of decision constants are shown. The hydropathy profile is calculated exactly as recommended by Kyte and Doolittle (3) using a window nine residues wide.

The sequence for human adenylate kinase whose structure is plotted in Figure 1 was taken from the entry called "Kihua" in the Protein Sequence Database of the Protein Identification Resource of the National Biomedical Research Foundation at Georgetown University.

### OUTPUT

#### Graphic Output

PEPLOT produces a plot that consists of 8 separate subplots or panels (labelled A - H in Figure 1). Each panel contains a statistic plotted versus a position in the sequence. To increase readability, the plots are drawn in four colors, but are reproduced here in black and white. Throughout panels B to G, dotted lines are used to indicate measures of alpha-helicity and solid lines are used for beta-sheets and turns. The hydropathy profile is also shown as a solid line.

A) Sequence and Sequence Schematic The amino acid sequence is written at the bottom of panel A at the same scale as the plots in the other panels. This allows immediate correlation of any features of interest with the protein sequence. The sequence schematic uses a variety of symbols to represent the amino acids of the sequence. Charged residues are shown as long dotted

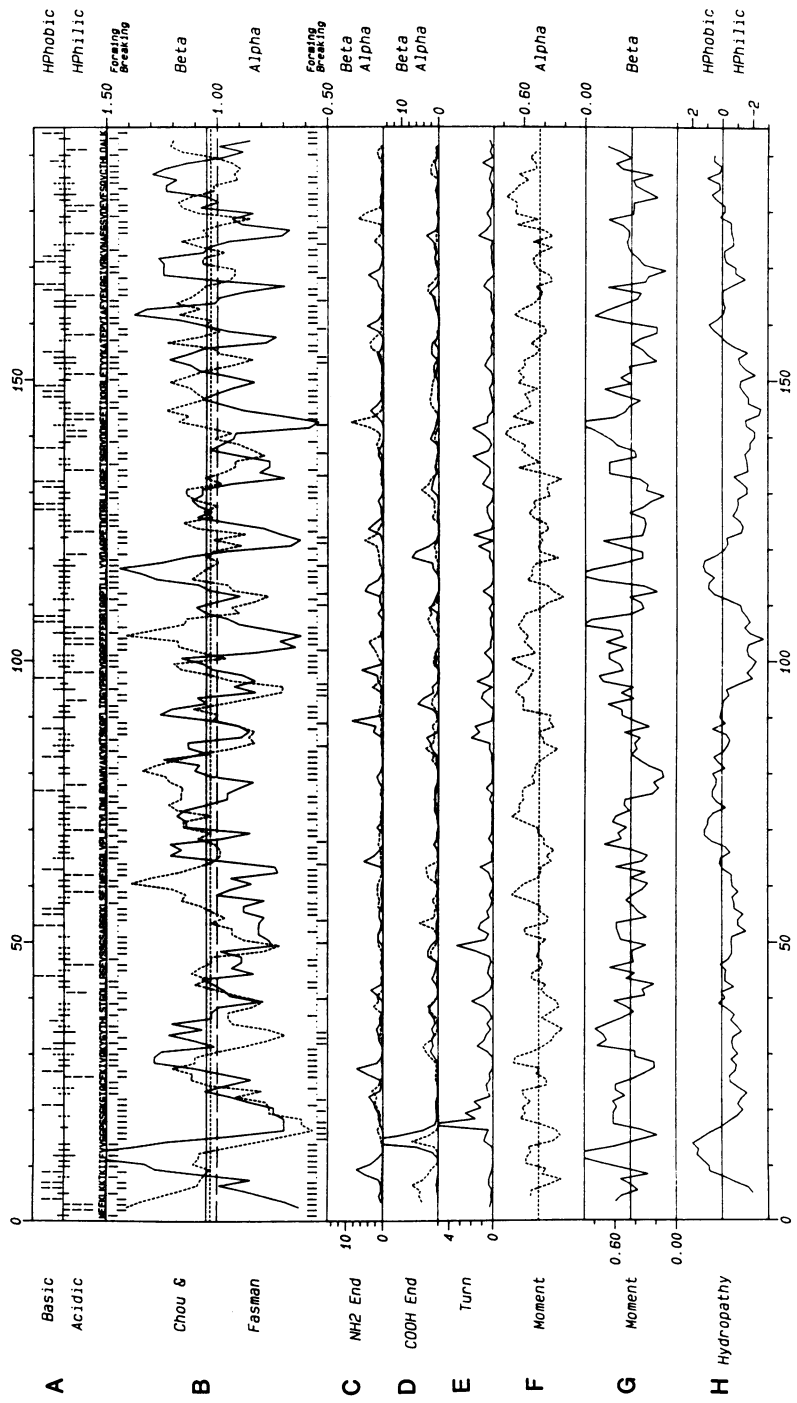


Figure 1 PEPLOT graphic output for human adenylate kinase. Dotted lines are used to distinguish the curves for alpha structures from the curves for beta structures.

## Nucleic Acids Research

---

(green) lines extending above (basic) and below (acidic) the central line. Differences in the length of the lines are used to distinguish between the different acidic and basic residues, with length increasing with alphabetic order. Hydrophobic residues are shown as solid (blue) lines crossing the central line. Two classes of hydrophobic residues are indicated by different line lengths: aliphatic residues (met, ile, leu, val) are shown as long lines, and aromatic residues (phe, trp, tyr) are shown as short lines. Hydrophilic residues are similarly indicated by dotted (red) lines crossing the central line, with gln and asn shown as long lines, and ser and thr as short lines. Pro is marked with a short (black) line, and ala, cys, and gly are unmarked. The symbol for pro is slightly shorter than the aromatic residue symbol to make them easier to distinguish on black and white plots.

B) Alpha and Beta Structure Potentials The Chou and Fasman alpha and beta conformational potentials (1, table V) are plotted as running averages of four adjacent residues. The horizontal dotted and solid lines indicate the minimum levels for predicting alpha and beta structures. Positions below the dashed line indicate possible tetrapeptide breakers. The small hashmarks at the top of this panel classify each residue as forming, breaking, or indifferent for beta structures. Likewise the hashmarks at the bottom of the panel classify each residue as alpha-forming breaking or indifferent.

C) and D) Alpha and Beta Structure End Predictions The Chou and Fasman methods provide a means of calculating the likelihood that a region of a sequence is at the amino or carboxyl end of an alpha or beta structure. Probabilities of each residue occurring at the end of alpha or beta structure, or in the region just beyond the end of the structure, were calculated by Chou and Fasman (1, Tables VI and VII). The product of the boundary conformational parameters for the three residues before and after the current position in the sequence are plotted for amino terminal ends (C) and carboxyl ends (D). These panels are most useful in resolving the structures of regions with fairly high potentials for both alpha and beta structures. A peak in any of the four curves indicates an increase in the potential for an end at that position.

E) Reverse Turn Potential The reverse turn potential is the product of each residue in a four residue window occurring at each position of the turn (1, table VIII). Chou and Fasman recommend using a minimum threshold of 1.5 to 2.0 times the average turn potential for prediction of reverse turns. This is equivalent to 0.81 to 1.08 on the scale shown in panel E.

F) and G) Hydrophobic Moment These panels show a measure of the hydrophobic moment for inter-residue angles corresponding to alpha and beta structures

(5). Each panel shows the maximum value of the hydrophobic moment within the range of angles characteristic of that structure. A classical alpha helix contains 3.4 residues per turn or an inter-residue angle of  $100^\circ$ . The hydrophobic moment shown in panel F is the maximum value within the range  $90^\circ$  to  $110^\circ$ , calculated over eight residues. Beta structures typically have inter-residue angles of  $160^\circ$  but the range of observed angles is broader than for alpha helices. For this reason we show the maximum hydrophobic moment for the range of  $140^\circ$  to  $180^\circ$  in panel G. In both cases, the hydrophobic moment is calculated for angular increments of  $1^\circ$  within the stated range. Segments of secondary structure that are located so that one side faces the hydrophobic interior of the protein and the other faces the exterior or solvent have high hydrophobic moments (i.e. greater than 0.4). The hydrophobic moment may also be useful for helping locate transmembrane segments which typically have low hydrophobic moments, and high hydrophobicities (6). These panels may therefore allow one to predict whether the secondary structures predicted by the Chou and Fasman or Robson and coworkers method are located on the surface of the protein, are transmembrane segments, or are within the globular core of the protein.

H) Hydropathy Profile The last panel shows a hydropathy profile; a running average over nine residues of the hydropathy scale of Kyte and Doolittle (3). Peaks in this panel indicate hydrophobic (interior) regions, and valleys hydrophilic (surface) regions. Globular proteins typically produce hydropathy plots with hydropathy values ranging from +2 to -2. Trans-membrane segments of proteins usually produce much higher hydropathy scores (typically in the +3 to +4 range).

#### Printed Output

Robson Prediction Secondary structure prediction using the method of Robson et al is displayed as shown in Table 1. The structure at each residue is predicted by summing the propensity parameters for alpha, beta, reverse turn, or random coil within a window of 17 residues around the point in the sequence where the structure is being predicted. The propensity parameters are position specific, that is, they vary both with the residue and its position in the 17 residue window. The residue in the center of the window is predicted to participate in the structure with the highest summed propensity.

Decision constants may be used to bias the prediction if there is information from measurements of circular dichroism about the amounts of alpha and beta structure. This is done by adding or subtracting a constant from the alpha and beta summed propensities before comparing the propensities for the

Table 1

Garnier et al Structural Prediction (Printed Output)

Structural composition for no decision constant: alpha = 33.5% beta = 23.2%

%Alpha % Beta	No DC	<20 <20	<20 20-50	<20 >50	20-50 <20	20-50 20-50	20-50 >50	>50 <20	>50 20-50
Pos									..
18	C	C	C	C	C	C	C	C	C
19	C	C	C	C	C	C	C	C	C
20	T	T	T	T	T	T	T	T	T
21	T	T	T	T	T	T	T	T	T
22	T	T	T	T	T	T	T	T	T
23	T	T	T	T	T	T	T	T	T
24	T	T	B	B	T	B	B	T	B
25	B	B	B	B	B	B	B	A	B
26	B	B	B	B	B	B	B	B	B
27	B	B	B	B	B	B	B	B	B
28	B	B	B	B	B	B	B	B	B
29	B	B	B	B	B	B	B	B	B
30	B	T	B	B	T	B	B	A	B
31	T	T	B	B	T	B	B	A	B
32	T	T	T	T	T	T	T	T	T
33	T	T	T	T	T	T	T	T	T

four structures and making the structural prediction. In the absence of structural information, Garnier et al (2) suggest using the percent alpha and beta structures determined using no decision constants to choose the decision constants for the final structural prediction. (The calculated percent alpha and beta structure using no decision constant is shown at the beginning of the table.) Predictions for all of the combinations of decision constants recommended by Garnier et al are shown in Table 1 as well as a prediction using no decision constant.

Chou and Fasman and Kyte and Doolittle The measurements used to create the Chou and Fasman and hydrophathy plots are also available in written form (Table 2). The residue, alpha and beta parameters, average alpha and beta potentials, end prediction, and reverse turn prediction are shown along with the average hydrophathy.

Hydrophobic Moment The maximum hydrophobic moment and the inter-residue angle producing the maximum can also be shown in written form (Table 3).

#### DISCUSSION

PEPLOT makes a display of the standard measurements of protein secondary structure on a single coordinated plot. Secondary structure prediction is subject to great uncertainty so the output of PEPLOT is designed to provide

Table 2  
Chou and Fasman Structural Measures (Printed Output)

Pos	Res	Alpha	Alpha	Beta	Beta	Alpha		Beta		Turn	HPhob
		Stat	Ave	Stat	Ave	NH2	COOH	NH2	COOH		
18	G	0.57	0.77	0.75	0.75	1.67	0.36	2.22	0.69	2.56	-1.23
19	S	0.77	0.77	0.75	0.75	2.21	0.73	2.73	0.65	1.12	-1.62
20	G	0.57	0.78	0.75	0.86	1.75	0.18	3.90	0.19	1.76	-1.03
21	K	1.16	0.92	0.74	0.95	3.05	0.22	0.87	0.14	0.30	-0.52
22	G	0.57	0.80	0.75	1.06	1.23	0.32	0.53	0.13	0.52	-0.48
23	T	0.83	1.04	1.19	0.96	0.57	0.78	1.08	0.20	0.63	-0.87
24	Q	1.11	1.12	1.10	0.85	0.14	0.70	4.41	0.10	0.29	-0.93
25	C	0.70	1.11	1.19	0.98	0.34	0.49	7.27	0.11	0.36	-0.59
26	E	1.51	1.20	0.37	1.10	0.40	0.74	1.65	0.20	0.04	-1.01
27	K	1.16	1.10	0.74	1.29	0.32	2.24	0.48	1.14	0.05	-0.70
28	I	1.08	1.10	1.60	1.29	0.15	3.51	0.18	2.19	0.07	-0.62
29	V	1.06	1.01	1.70	1.25	0.20	4.22	0.36	1.48	0.55	-0.70
30	Q	1.11	0.88	1.10	1.02	0.42	2.21	1.15	1.86	1.47	-1.26
31	K	1.16	0.78	0.74	1.11	0.76	0.86	1.32	1.21	0.85	-0.94
32	Y	0.69	0.70	1.47	1.22	0.63	0.30	1.81	1.87	0.63	-0.56
33	G	0.57	0.77	0.75	1.07	0.48	0.33	0.87	0.92	0.23	-0.80

all of the information from each of several different kinds of measurements. We feel that this display is more informative than techniques that predict secondary structure by making arbitrary choices that cast a shadow over dissonant information. PEPLOT leaves the analysis and structural prediction, if possible, to the user.

Table 3  
Hydrophobic Moment (Printed Output)

Pos	Res	Alpha	Alpha	Beta	Beta
		Angle	Value	Angle	Value
18	G	111.0	0.53	181.0	0.59
19	S	111.0	0.60	181.0	0.63
20	G	111.0	0.62	181.0	0.50
21	K	111.0	0.38	181.0	0.60
22	G	90.0	0.38	181.0	0.60
23	T	107.0	0.48	140.0	0.43
24	Q	101.0	0.45	140.0	0.32
25	C	92.0	0.70	180.0	0.22
26	E	96.0	0.68	181.0	0.22
27	K	110.0	0.51	180.0	0.41
28	I	110.0	0.41	180.0	0.41
29	V	110.0	0.33	141.0	0.78
30	Q	110.0	0.30	158.0	0.69
31	K	111.0	0.22	171.0	0.72
32	Y	110.0	0.32	141.0	0.80
33	G	110.0	0.42	140.0	0.67

We have found that plots of structural measurements like PEPLOT can be of significant value for detecting similarity between sequences. Similarity that is difficult for the human or the computer to recognize in the primary structure may generate patterns of secondary structure measurements that are recognizably similar.

We do not believe that the state-of-the-art allows secondary prediction for most protein molecules based on primary structure alone. Tools like PEPLOT are useful, however, for identifying small elements of protein structure. For instance, the hydrophobic moment and hydrophathy sections of the plot may be useful for determining which segments of a protein are located at the surface of the protein. Many scientists are interested in predicting antigenic sites on proteins. The most antigenic active sites seem to be located on the surface of the protein, but single measures such as hydrophathy or the presence of reverse turns have been only moderately successful in predicting their locations (4,8). The several kinds of measurements shown by PEPLOT may help make these kinds of predictions.

### ACKNOWLEDGEMENTS

PEPLOT was developed in the UW Biotechnology Center with funds provided by the University of Wisconsin Genetics Computer Group (UWGCG). UWGCG is a consortium of institutions desiring VAX/VMS software for biological sequence analysis. All fees paid by consortium members are used for the development of computer applications for the field of Molecular Genetics.

### REFERENCES

1. Chou, P.Y. and Fasman, G.D. (1978) *Adv. Enz.* 47, 45-147.
2. Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) *J. Mol. Biol.* 120, 97-120.
3. Kyte, J. and Doolittle, R.F. (1982) *J. Mol. Biol.* 157, 105-132.
4. Hopp, T.P. and Woods, K.R. (1981) *Proc. Natl. Acad. Sci.* 78, 3824-3828.
5. Eisenberg, D., Sweet, R.M. and Terwilliger, T.C. (1984) *Proc. Natl. Acad. Sci.* 81, 140-144.
6. Devereux, J., Haerberli, P. and Smithies, O. (1984) *Nucl. Acids Res.* 12, 387-395.
7. Eisenberg, D., Weiss, R.M. and Terwilliger, T.C. (1982) *Nature* 299, 371-374.
8. Westhof, E., Altschuh, D., Moras, D., Bloomer, A.C., Mondragon, A., Klug, A. and Van Regenmortel, M.H.V. (1984) *Nature* 311, 123-126.