Hopp, T.P. (1987) Improved hydrophilicity plotting method for membrane proteins.
In: Methods in protein sequence analysis 1986, pp. 601-607. Walsh, K.A., ed.
Humana Press, Clifton, New Jersey. http://thomas-hopp.com/THscience2a.html

IMPROVED HYDROPHILICITY PLOTTING METHOD FOR MEMBRANE PROTEINS

Thomas P. Hopp

Immunex Corporation
51 University St., Seattle, Washington 98101

SIX POINTS TO REMEMBER CONCERNING HYDROPHILICITY, HYDROPHOBICITY, AND HYDROPATHY PLOTS

1. Most methods give similar results. Figure 1 compares plots obtained using the various scales currently available. It is readily apparent that all of the plots give much the same information about the protein, because all of the major upward and downward deflections are common to all plots, and most of the minor deflections are shared as well.

2. Most scales readily identify membrane spanning and signal sequences. Fig. 1 also demonstrates that signal and transmembrane segments appear as wide valleys (usually about 20 residues across) and can be distinguished from valleys corresponding to secondary structure elements (α helices and β strands), which range from 6 to 10 residues across. Narrower valleys correlate well with buried segments of helices and strands, and peaks correlate with highly exposed loops of peptide chain (12,13). This relationship is true for all profiles, but is optimal with the acrophilicity scale (12).

3. The most significant differences between scales are in their hydrophilic ends. Proper placement of the hydrophilic amino acids has a strong effect on prediction of secondary structures and protein interaction sites (Table 1 and ref. 12). Placement of the four highly charged amino acids (Asp, Glu, Arg, Lys) at the top of the scale with equivalent values enhances predictive ability for protein surface interaction sites such as antigenic

determinants, limited proteolysis sites, post-translational modification sites, protein-protein binding sites and protein-DNA or protein-RNA contact sites (12). Equivalence of these values, as in the original Hopp and Woods scale (1), eliminates bias toward either positively or negatively charged sites. All scales other than the Hopp and Woods scale have such a built-in bias. On the other hand, proper placement of the *small* amino acids (Ser, Pro, Asn, and especially Gly) has a great influence on correlation with secondary and tertiary structure.

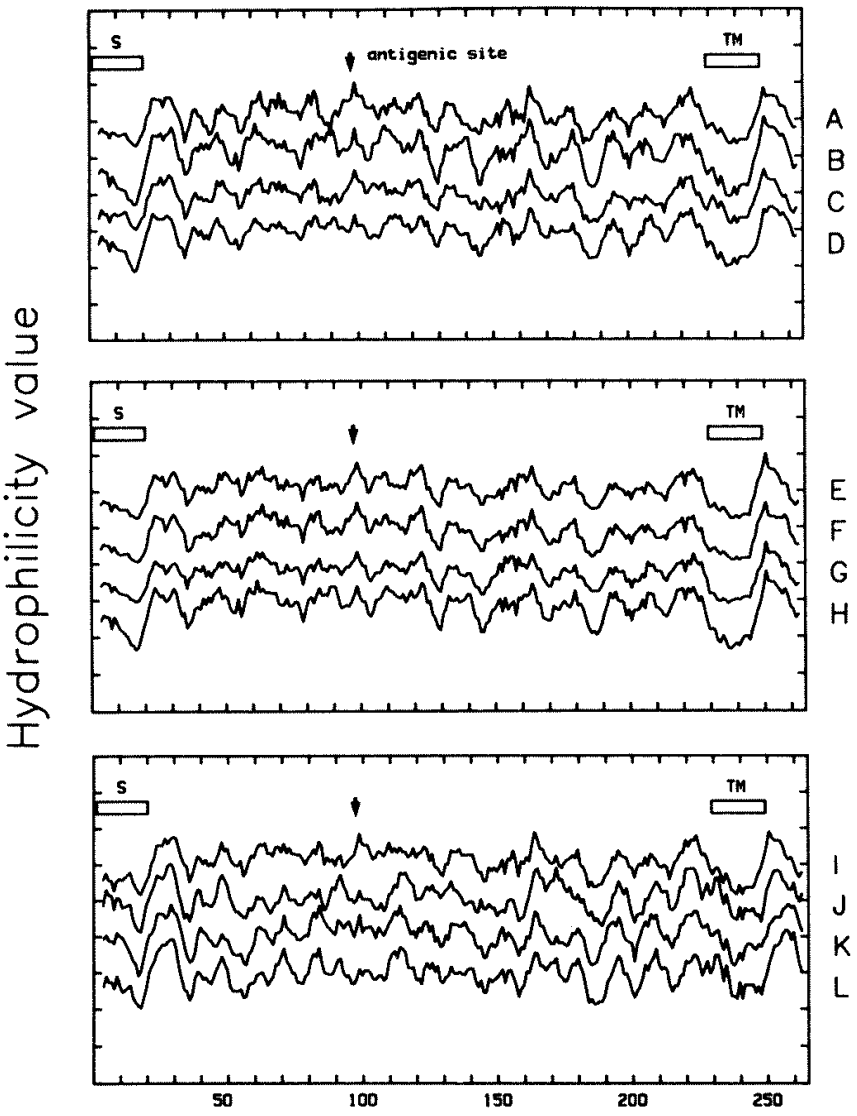

Figure 1. Comparison of hydrophilicity/-phobicity plots for the IAβ^d histocompatibility antigen. S, signal peptide; TM, transmembrane anchor segment. The antigenic site is a point mutation that causes a shift in T-cell antigenicity (15). Scales used to generate the profiles are: A, Hopp and Woods (1); B, Kyte and Doolittle (2); C, Eisenberg and McLachlan (3); D, Rose et al. (4); E, Sweet and Eisenberg (5); F, von Heijne (6); G, Janin (7); H, Chothia (8); I, Fauchere and Pliska (9); J, Karplus and Schulz (10); K, Chou and Fasman (strands) (11); L, Acrophilicity (12).

4. Averaging group length is critical. An optimum correlation of hydrophobic valleys with secondary structures (α-helices, β-strands) is achieved with an averaging-group length of six (Fig. 2 & Table 1). Surface features such as antigenic sites are best identified with hexapeptide averages, as well (1). That these two independent criteria both implicate six-residue averages as optimal, suggests that hexapeptide averaging should be considered the standard procedure in hydrophilicity/-phobicity plotting. The second best results are obtained with pentapeptide averages, in both cases.

Table 1. Optimum Windows for Correlation of Valleys With α-Helices and β-Strands. The procedure used to generate Fig. 4 was applied to all available hydrophilicity/-phobicity scales. The averaging group length that performed best is listed under "optimum window", along with the rate at which it located secondary structure elements.

	Optimum Window		Second Best Window		Antigenic Sites
Scale	Window Length	Success Rate(%)	Window Length	Success Rate(%)	Percent Correct
Hopp and Woods	6	64.6	5	62.7	88
Kyte and Doolittle	5/6	68.7	5/6	68.7	65
Acrophil-icity	6	74.3	5/7	73.5	59
Rose et al	4	67.5	5	66.6	68
Sweet and Eisenberg	6	64.6	5	64.5	56
Von Heijne	6	64.4	5	63.6	64
Chou and Fasman					
(turns)	7	78.4	6	78.1	43
(helices)	8	73.8	6	72.8	68
(strands)	6	70.6	5	70.2	N.D.[c]
Fauchere & Pliska	6	66.0	4	65.4	N.D.
Chothia	6	66.1	5	65.5	60
Janin	4	60.7	5	59.8	68
Random	10/2	54.1	6	53.9	57
Concensus[b]	6	–	5	–	–

[a]From Table 1, Column 3, Ref. 1.
[b]Not including random.
[c]Not determined.

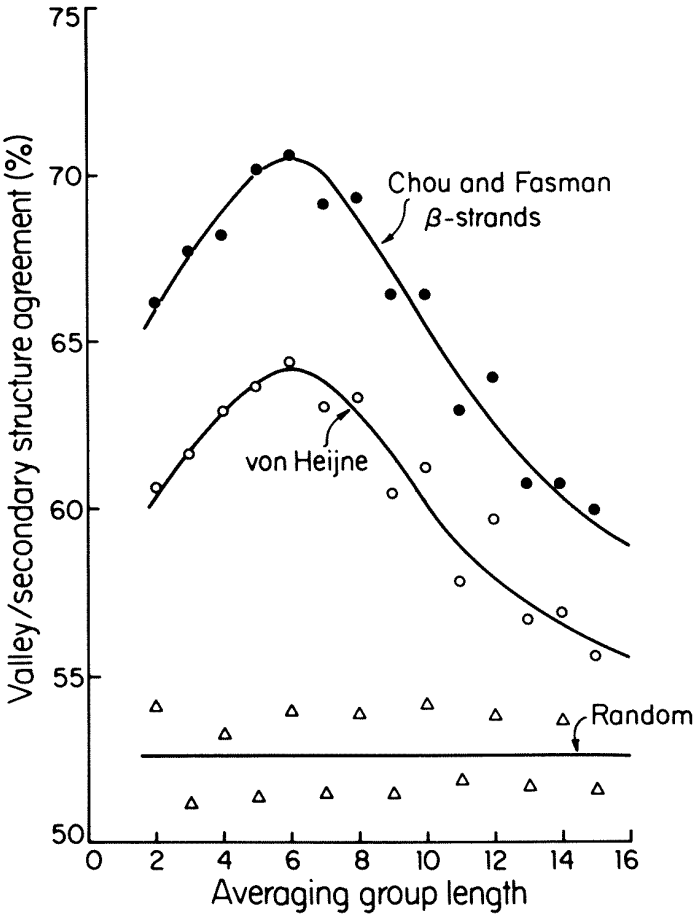

Figure 2. Correlation of valleys with secondary structures. Using a data set of 70 proteins for which 3-D structures are known, the occurrence of the lowest 20% of points on hydrophilicity plots was correlated with the presence of α-helix or β-strand. This process was repeated for all averaging group lengths from 2 to 15.

The earliest hydrophilicity/-phobicity plotting publications recommended windows of 5 (13) and 6 (1), but this has largely been ignored, perhaps because Kyte and Doolittle erroneously stated that longer windows are more useful, in their more recent publication (2). Their statements are untenable however, because they were made without the benefit of experimental evidence. As Table 1 shows, *most* methods work best with hexapeptide averages, *including* that of Kyte and Doolittle.

5. Improvements can be made using older scales with new computer methods. The earliest hydrophilicity/-phobicity plots (13) were made using values derived from Nozaki and Tanford's amino-acid hydrophobicity measurements (14), which did not assign values to any of the hydrophilic amino acids. Hopp and Woods (1) were the

first to utilize a full scale that included values for the hydrophilic amino acids, based on an extrapolation of the work of Nozaki and Tanford. Although many scales have been devised since then, none represent a real improvement over the original Hopp and Woods full scale of amino-acid values. However, hydrophilicity plots can be improved by other means. For example, a new computer algorithm has been published (HYDRO3; Ref. 12) that recognizes particular Gly, Ser, or Thr residues that are likely to be buried in protein interiors or in transmembrane segments. As can be seen in Fig. 4, this procedure can be used to enhance the appearance of transmembrane segments without disturbing the distribution of the most hydrophilic peaks.

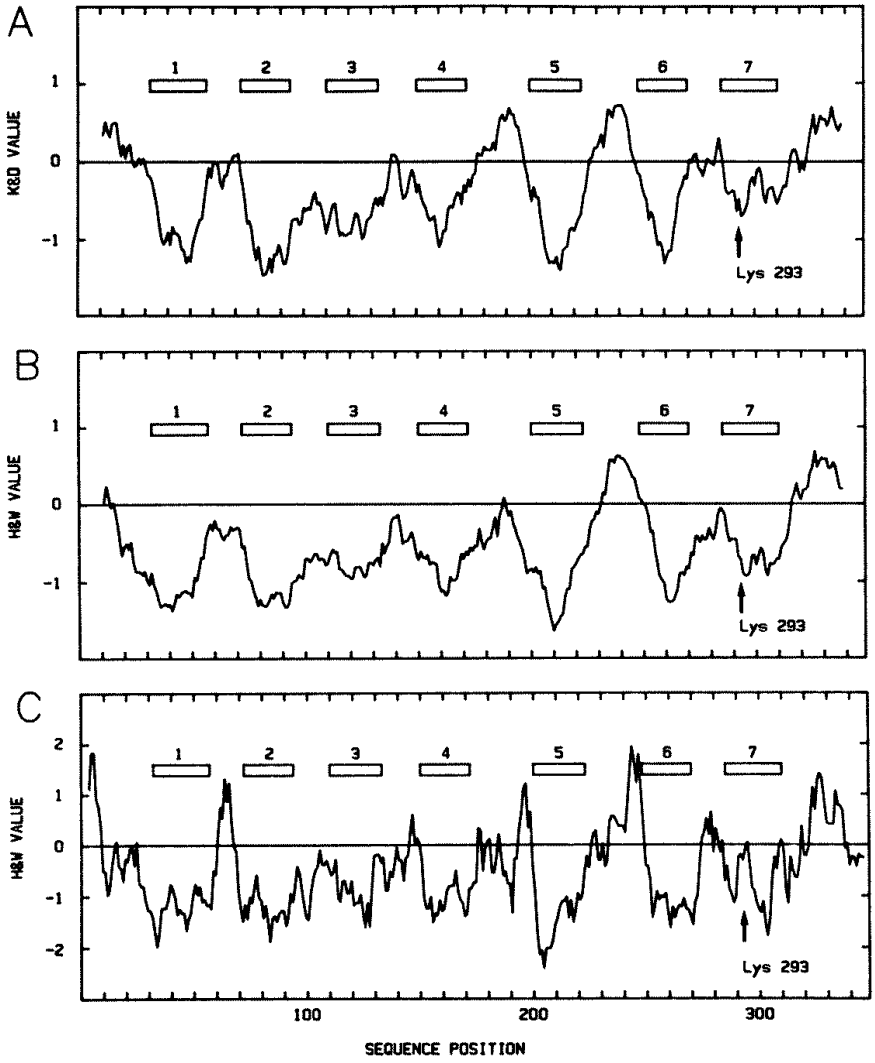

Figure 3. Hydrophilicity plots for human blue visual pigment. Panels A and B were made using 21-residue windows and the Kyte and Doolittle (A) or Hopp and Woods (B) scales. Panel C used the Hopp and Woods scale with a six-residue window and the Gly, Ser and Thr adjustments of the HYDRO3 procedure. Bars mark the seven membrane-associated regions; Lys 293 is derivatized with retinal.

6. Hydrophilicity plotting can be optimized for extracting maximum information from protein sequences. Using six-residue averages, an appropriate scale, and updated computer algorithms leads to hydrophilicity plots that yield more information. As an example, Fig. 3C shows a hexapeptide hydrophilicity plot, using HYDRO3 and the Hopp and Woods amino acid scale to analyze the structure of human blue visual pigment. This plot readily shows the seven putative transmembrane segments that appear as valleys in the two plots above, but also shows considerable other detailed information. Most notably, it can be seen that the retinal binding lysine 293 lies in a moderately hydrophilic segment of peptide chain, not in a hydrophobic segment, as the two other low resolution plots imply. Furthermore, it appears from the narrow valleys in this region of the profile that segment 7 may actually comprise at least three separate segments of secondary structure, either short α-helical segments or β-strands.

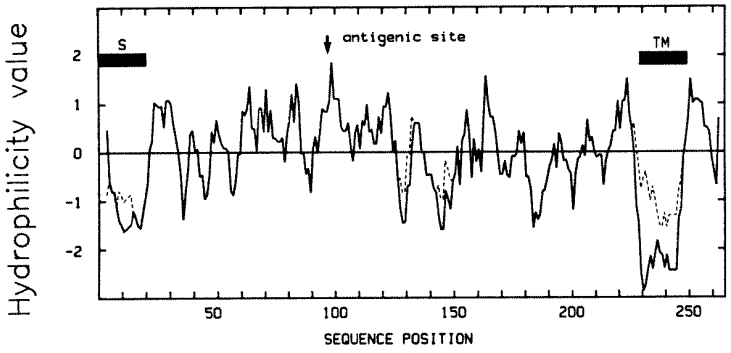

Figure 4. Demonstration of HYDRO3 effects for the IAβ^d histocompatibility antigen, The dotted profile represents the original Hopp and Woods method. The solid profile is from the HYDRO3 procedure. Lowering selected Gly, Ser and Thr values has enhanced the appearance of the signal (S) and transmembrane (TM) regions without interfering with the identification of the antigenic site as the highest peak on the profile.

REFERENCES

Hopp, T.P. & K.R. Woods, 1981, P.N.A.S. USA 78, 3824.
Kyte, J, & R.F. Doolittle, 1982, J. Mol. Biol. 157, 105.
Eisenberg, D, & A.D. McLachlan, 1986, Nature, 319, 199.
Rose, G.D., A.R. Geselowitz, G.J. Lesser, R.H. Lee & M.H. Zehfus, 1985, Science, 229, 324.
Sweet, R.M., & D. Eisenberg, 1983, J. Mol. Biol. 171, 479.
Von Heijne, G., 1981, Eur. J. Biochem. 116, 419.
Janin, J., 1979, Nature (London) 277, 491.
Chothia, C., 1976, J. Mol. Biol. 105, 1.
Fauchere, J.L., & V. Pliska, 1983, Eur. J. Med. Chem. Chim. Ther. 18, 369.
Karplus, P.A., & G.E. Schulz, 1985, Naturviss. 72, 212.
Chou, P.Y. & G.D. Fasman, 1978, Adv. Enzymol. 47, 45.
Hopp, T.P., 1986, J, Immunol. Methods, 88, 1.
Rose, G.D., 1978, Nature, 272, 586.
Nozaki, Y. & C. Tanford, 1971, J. Biol. Chem. 246, 2211.
Brown, et. al. 1986, Science, 231, 255.