Computer programs to identify and classify amphipathic α helical domains

Martin K. Jones, G. M. Anantharamaiah, and Jere P. Segrest[1]

Atherosclerosis Research Unit, Departments of Medicine and Biochemistry, 630 DERB, UAB Medical Center, Birmingham, AL 35294

Abstract The amphipathic α helix is an often-encountered secondary structural motif in biologically active peptides and proteins. An amphipathic helix is defined as an α helix with opposing polar and nonpolar faces oriented along the long axis of the helix. In a recent review article we grouped amphipathic helixes into seven distinct classes (A, H, L, G, K, C, and M) based upon a detailed analysis of their physical–chemical and structural properties (Segrest, J. P., et al. Amphipathic helix motif: classes and properties. *Proteins*. 1990. **8**: 103–117). We have developed five computer programs that automate analysis and classification of potential amphipathic helical domains from primary amino acid sequence data. Here we describe these five programs and illustrate their usefulness by comparing two data sets of sequences representing different amphipathic α helical motifs from the exchangeable apolipoproteins. In a companion review article (Segrest, J. P., et al. The amphipathic helix in the exchangeable apolipoproteins: a review of secondary structure and function. *J. Lipid Res.* 1992. **33**: 000–000) these five programs are used to localize and characterize the putative amphipathic helixes in the exchangeable apolipoproteins.—**Jones, M. K., G. M. Anantharamaiah, and J. P. Segrest.** Computer programs to identify and classify amphipathic α helical domains. *J. Lipid Res.* 1992. **33**: 287–296.

Supplementary key words protein–lipid interactions • computer prediction of protein structure • protein folding • protein motifs • snorkel hypothesis • hydrophobic moment • helical wheel • helical net

Amphipathic helical domains have been reported in a variety of proteins (1, 2). The functional properties suggested for amphipathic helixes include lipid association, membrane perturbation in the form of fusion or lysis, hormone-receptor catalysis, transmembrane signal transduction, regulation of kinase-calmodulin signal transduction and transmembrane helical bundle formation. In a recent review article (2) we grouped amphipathic helixes into seven distinct classes (A, H, L, G, K, C, and M) based upon a detailed analysis of their physical–chemical and structural properties. Differences in polar face charge distribution dominated the differences between the classes. Here we describe five computer programs (WHEEL, HELNET, COMBO, COMNET, and CONSENSUS) we have developed to automate amphipathic helix identification and classification.

METHODS

Amino acid sequence database

Amino acid sequences for the apolipoproteins analyzed in this study were obtained from the National Biomedical Research Foundation database. Selection of amphipathic helical domains for inclusion into each of the two data sets examined here was performed as described in a companion review article (3).

Computer programs for analysis of amphipathic helixes

Helical wheel program (WHEEL). The WHEEL program creates a "Schiffer-Edmundson" helical wheel diagram (4) of a given sequence of amino acids (up to 36 in number) arranged as an ideal α helix (100° rotation per residue) seen down the long axis from the amino terminal end. The residues are projected onto a circular figure denoting the helix that shows each amino acid residue with charge and sequence number. Optionally, each residue can be shaded according to its hydrophobicity. The circular figure is rotated so as to orient the hydrophobic face toward the top of the page. There are two separate options to determine this orientation.

The default option uses the hydrophobic moment ($<\mu_H>$) originally devised by Eisenberg, Weiss, and Terwilliger (5) to orient the hydrophobic face. **Fig. 1** diagrams the basic elements of the hydrophobic moment concept (not drawn to scale). A hydrophobicity scale is used to assign a vector magnitude to each amino acid residue around a helical wheel. In the studies reported here we have used a normalized GES

[1]To whom correspondence should be addressed.

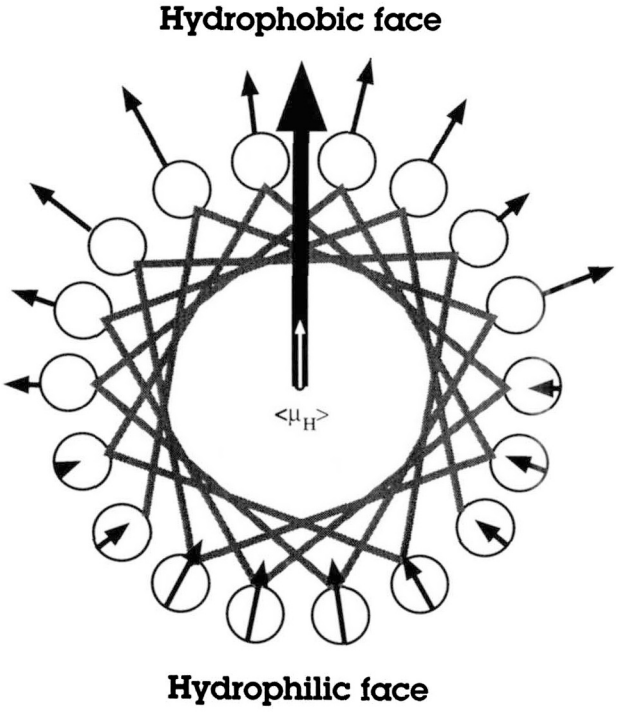

Fig. 1. Schematic helical wheel diagram to illustrate the basic features of the hydrophobic moment for analysis of amphipathic helixes. The long axis of the idealized α helix is perpendicular to the plane of the page. Small black arrows, vectors representing the hydrophobicity of each amino acid residue on the wheel—if hydrophobicity is positive, the vector points radially outward from the wheel center and if negative, the vector points inward toward the wheel center; large black arrow, vector representing the vector sum (total hydrophobic moment) of the individual amino acid residue vectors; small white arrow, mean hydrophobic moment per residue.

hydrophobicity scale (6) but any hydrophobicity scale can be used. The values we used are: Phe, 1.036; Met, 0.975; Ile, 0.913; Leu, 0.852; Val, 0.811; Trp, 0.668; Cys, 0.689; Ala, 0.607; Thr, 0.525; Gly, 0.484; Ser, 0.402; His, 0.333; Pro, 0.239; Tyr, 0.137; Orn, 0.000; Gln, –0.558; Asn, –0.701; Glu, –1.396; Lys, –1.518; Asp, –1.600; Arg, –2.233.

For residues with a positive hydrophobicity, the vectors (small black arrows) are directed radially away from the wheel center; for residues with a negative hydrophobicity (i.e., hydrophilic residues), the vectors are directed radially toward the wheel center. The vectors for all residues in a given wheel are summed and the vector sum (large arrow) is the net hydrophobic moment. Since the net hydrophobic moment is dependent on the length of the helix analyzed, the mean hydrophobic moment per residue is used (small white arrow). In the default option, the wheel is oriented so that the hydrophobic moment is directed toward and perpendicular to the top of the page.

By specifying the second option, WHEEL/SNORKEL, the wheel orientation to the page is realigned so that the normal to the top of the page bisects the nearest positive residues to either side of the hydrophobic moment. The rationale behind this method of orientation is schematically illustrated in **Fig. 2** (not drawn to scale) and is as follows:

The most distinctive feature of the amphipathic helix motif called class A, as defined in our previous review (2), is the unique clustering of positively charged amino acid residues at the polar–nonpolar interface and negatively charged residues at the center of the polar face. Using peptide analogs of amphipathic helixes of the A class we have shown that this clustering motif is important for lipid affinity (7–12). Note that the bulk of the van der Waals' surface areas of the positively charged residues are hydrophobic. It is proposed that these amphipathic basic residues, when associated with phospholipid, extend ("snorkel") toward the polar face of the helix to insert their charged moieties into the aqueous milieu (Fig. 2). Thus, essentially the entirety of the uncharged van der Waals' surface of the amphipathic helixes of the apolipoproteins can be buried within the hydrophobic interior of a phospholipid monolayer. The snorkel hypothesis predicts, therefore, that the position of Lys and Arg residues, rather than the hydrophobic moment, would tend to dominate the radial orientation of class A amphipathic helixes relative to the lipid surface.

Helical net program (HELNET). This program creates a diagram by the method of Lim (13) of the α helix seen as a cylinder cut along the center of the polar face and flattened. The center of the hydrophobic face (dotted line), determined by the hydrophobic moment, lies in the center of the figure and is oriented to rise out of the page. HELNET/SNORKEL is also available to realign the center of the hydrophobic face as in WHEEL/SNORKEL. An additional program option, HELNET/ANGLE, can change the angle at which the cylinder is cut relative to the line running down the center of the polar face. The approximate angle can be determined by observation of the three reference

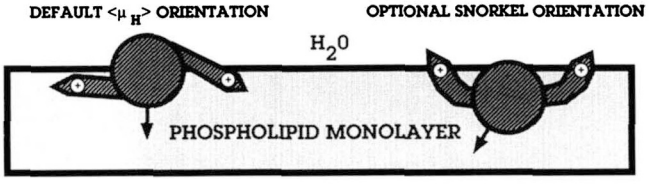

Fig. 2. Schematic representation of the physical chemical rationale for the default and snorkel options for orientation of helical wheels. Each circle represents the backbone of two identical amphipathic α helixes whose long axes are directed outward from the page. The arrows represent the hydrophobic moment for each helix. The arms attached to each helix represent the two positively charged amino acid residues that are on opposite sides of the moment vector and nearest the center of the nonpolar face.

lines (single dotted line for the center of the nonpolar face and two dashed lines for the polar–nonpolar interfacial planes).

Program for addition of helical wheels (COMBO). This program superimposes and averages a set of wheels. Before the wheels are superimposed, each wheel is rotated so that the hydrophobic face points towards the top of the page (**Fig. 3A**). Selected residues of the set are projected onto two wheel diagrams. The left-hand wheel displays the counts of all positively charged residues and the right-hand wheel displays the counts of all negatively charged residues.

The input to the program is a list of amino acid sequences identical to those accepted by the helical wheel program. By specifying a program option, COMBO/SNORKEL, the hydrophobic moment vector for each wheel is realigned as in WHEEL/SNORKEL.

The wheel for the positive residues displays at every 10-degree position around the helix the count of positively charged residues that fall into that position. On the hydrophobic face the average hydrophobicities of the uncharged residues are displayed at every 20-degree position. The wheel for the negative residues displays at every 10-degree position around the helix the count of negatively charged residues that fall into that position and the total number of residues for the set of sequences that occur at every 20-degree position. The option COMBO/SELECT = AA1-AA2, where AA1 and AA2 are three-letter amino acid codes, changes from the default examination of positive and negative residues to the examination of two specific amino acids.

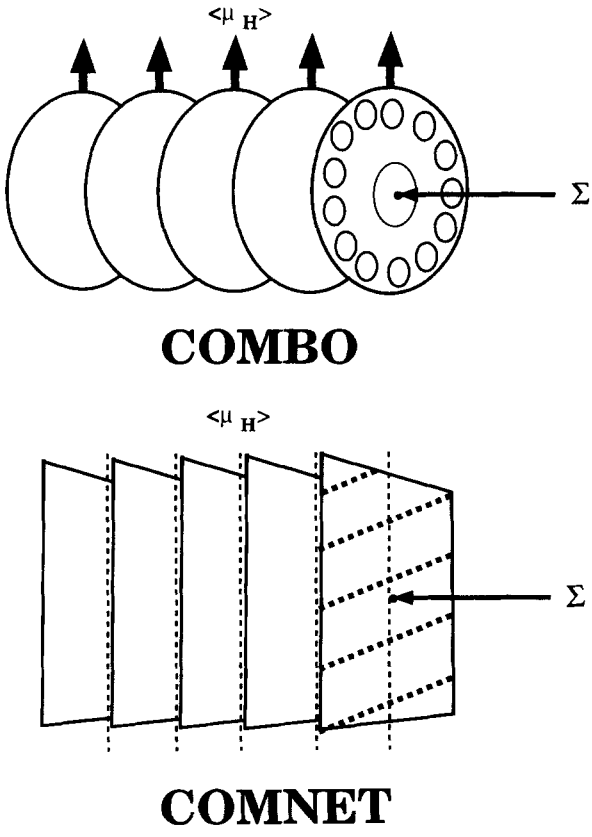

Fig. 3. Schematic representations of the COMBO and COMNET algorithms. A. COMBO; the arrows represent the aligned hydrophobic moments of a series of helical wheels. The Σ indicates that each aligned helical wheel in a given data set is summed. B. COMNET; vertical dotted lines represent the aligned hydrophobic moments of a series of helical nets. The Σ indicates that each aligned helical net in a given data set is summed.

The program also provides the following (not shown in the figures included in this paper): the list of file names containing each amino acid sequence, and, for the sum of the complete set of amphipathic helixes, the mean hydrophobic moment per residue, the mean hydrophobic moment per residue of the central six residue positions of the nonpolar face, the mean number of charged residues per 11-mer for positive, negative, and total charged residues, and the ratio of Lys to Arg residues.

Program for addition of helical nets (COMNET). This program superimposes and averages a set of helical nets. The nets are superimposed so that the midpoint of each net coincides (Fig. 3B). By specifying a program option, COMNET/SNORKEL, the hydrophobic vector of each net is realigned as in HELNET/SNORKEL. As in COMBO and COMBO/SELECT, selected residues of the set (either positive and negative, or two specific amino acids) are projected onto two helical net diagrams. By default, the longest helix (indicated by horizontal dotted lines) is used to scale the figure. However, COMBO/NET = LENGTH can be used to specify an exact length for comparison with other diagrams. An option /ANGLE can be used to rotate the net as in HELNET. Its default value in COMBO/NET is 180° so as to show the hydrophilic face in the center of the net.

Consensus wheel program (CONSENSUS). This program superimposes a set of helical wheels in the same fashion as COMBO and a single figure classifies the amino acid residues into five physical–chemical groups: positive (Arg, Lys), negative (Glu, Asp), polar (Asn, Gln), neutral (Tyr, Pro, His, Ser, Gly, Thr, Ala), and hydrophobic (Cys, Trp, Val, Leu, Ile, Met, Phe). CONSENSUS uses a graduated shaded contour to plot, at 20-degree intervals, the proportionally scaled radial distribution of these five classes of amino acid residues. This distribution was collected at 10-degree intervals and smoothed by splitting the oddly spaced values between the two adjacent evenly spaced values. Also a consensus amino acid residue appears for each 20-degree position if there is an amino acid residue

that occurs at that position most often and at least one-third of the time. A program option, CONSENSUS/SNORKEL, is also available.

RESULTS AND DISCUSSION

WHEEL analyses

Helical wheel analyses using WHEEL/SNORKEL are shown in **Fig. 4A** and **Fig. 5A** for apolipoprotein A-I [165–186] and apolipoprotein A-I [99–120], respectively. Examples using the default WHEEL program are shown in Figs. 4B and 5B for the same two sequences.

Apolipoprotein A-I [165–186] is a good example of the class A amphipathic helix motif. Note the clusters of positive charges at the polar–nonpolar interface and negative residues in the center of the polar face. For this particular sequence, the SNORKEL option has no effect on helix orientation. This amphipathic helix has a mean hydrophobic moment of 0.23 per residue.

Apolipoprotein A-I [99–120] is an example of another type of amphipathic helix motif called class Y (3). Typical of this class are two negatively charged residue clusters on the polar face separating the two arms and the base of the Y motif formed by three positively charged residue clusters. For this sequence, the SNORKEL option produces a 10° rotation in helix orientation. This amphipathic helix has a mean hydrophobic moment of 0.41 per residue.

COMBO analyses

Analyses using COMBO/SNORKEL are shown in Figs. 4C and 5C for class A and class Y amphipathic helical data sets, respectively. The class A data set represents a total of 7 sequences from apoC-I, C-II, C-III and A-II; the class Y data set represents a total of 12 sequences from apoA-IV and A-I. In these two figures the positive and negative residue clusters defining class A and class Y amphipathic helix motifs are readily apparent.

Figs. 4D and 5D represent COMBO/SNORKEL analyses of the class A and class Y data set, respectively, in which the SELECT option has been used to graphically display the radial distribution of the hydrophobic residues Leu and Val. For both data sets these amino acid residues are almost entirely confined to the nonpolar faces of the wheel display.

CONSENSUS analyses

Analyses using CONSENSUS/SNORKEL are shown in Figs. 4E and 5E for class A and class Y amphipathic helical data sets, respectively. This program provides a graphic summary of many of the average properties of any given amphipathic helix data set. A comparison of these two figures with the COMBO/SNORKEL analyses shown in Figs. 4C and 5C serves as a point of reference for the contour plot display by the CONSENSUS program of the radial distribution of positive and negative amino acid residues. This comparison also provides orientation for the contour plot display of the radial distribution of hydrophobic, neutral and polar residues by CONSENSUS. Note that the neutral residues tend toward a uniform radial distribution with no preference for the polar or nonpolar faces.

Additional information on prevalence of specific amino acid residues is provided by the CONSENSUS program. For example, the display that determines a possible consensus (radial) amino acid residue composition indicates that: *a*) for the class A motif (Fig. 4E), the positively charged clusters are both largely Lys and the single negatively charged cluster is primarily Glu; *b*) for the class Y motif (Fig. 5E), the hydrophobic face is greatly enriched in Leu and the two negative charge clusters are primarily Glu.

HELNET analyses

Helical net analyses centered on the polar face (HELNET/SNORKEL/ANGLE = 180) are shown in **Fig. 6A** and **Fig. 7A** for apoA-I [165–186] (class A) and apoA-I [99–120] (class Y), respectively. Hydrophobic residues have been shaded. These same figures, rotated 180° by the ANGLE program option to center on the nonpolar face, are shown in Figs. 6B and 7B. HELNET plots have the advantage of graphically displaying both the linear and radial distribution of individual amino acid residues. For both these examples positive residues cluster linearly along the polar–nonpolar face (solid lines). In Fig. 7A, Lys106, representing the base of the Y motif, can be seen in the middle of the polar face; the center of the class A example contains no positive residues (Fig. 6A).

COMNET analyses

COMNET/SNORKEL analyses of the class A and class Y data sets are shown in Figs. 6C and 7C centered on the polar face and are shown in Figs. 6D and 7D centered on the nonpolar face, respectively. The charge clustering characteristic for both class A and class Y amphipathic helixes can be seen to extend linearly along their length, creating in effect charged edges.

Algorithm for cluster quantification

The COMBO program contains an algorithm that analyzes clustering of residues in the positive and negative combo wheel diagrams; filled circle and error bars indicate the average angle and its standard deviation, respectively, subtended by each residue cluster.

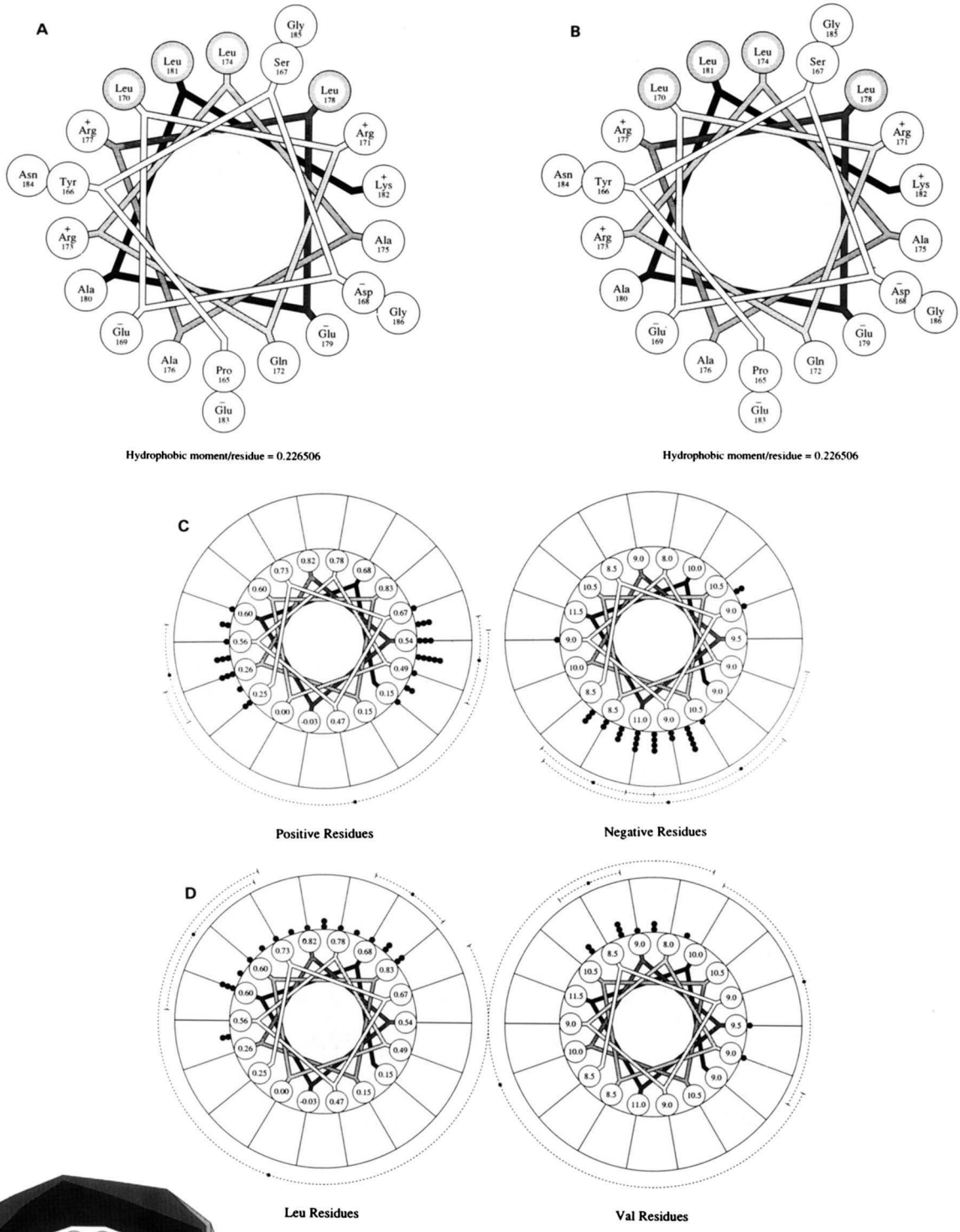

Fig. 4. Analyses of single and multiple class A amphipathic helixes using WHEEL, COMBO and CONSENSUS. A. WHEEL/SNORKEL analysis of apoA-I [165–186] that represents a class A amphipathic helix. B. Default WHEEL analysis of apoA-I [165–186]. C. COMBO/SNORKEL analysis of the seven class A amphipathic helixes from apoC-I [7–32, 33–53], apoC-II [14–39, 43–55], apoC-III [40–67], and apoA-II [7–30, 39–50, 51–71]. D. COMBO/SNORKEL/SELECT = LEU-VAL analysis of the seven class A amphipathic helixes. E. CONSENSUS/SNORKEL analysis of the seven class A amphipathic helixes.

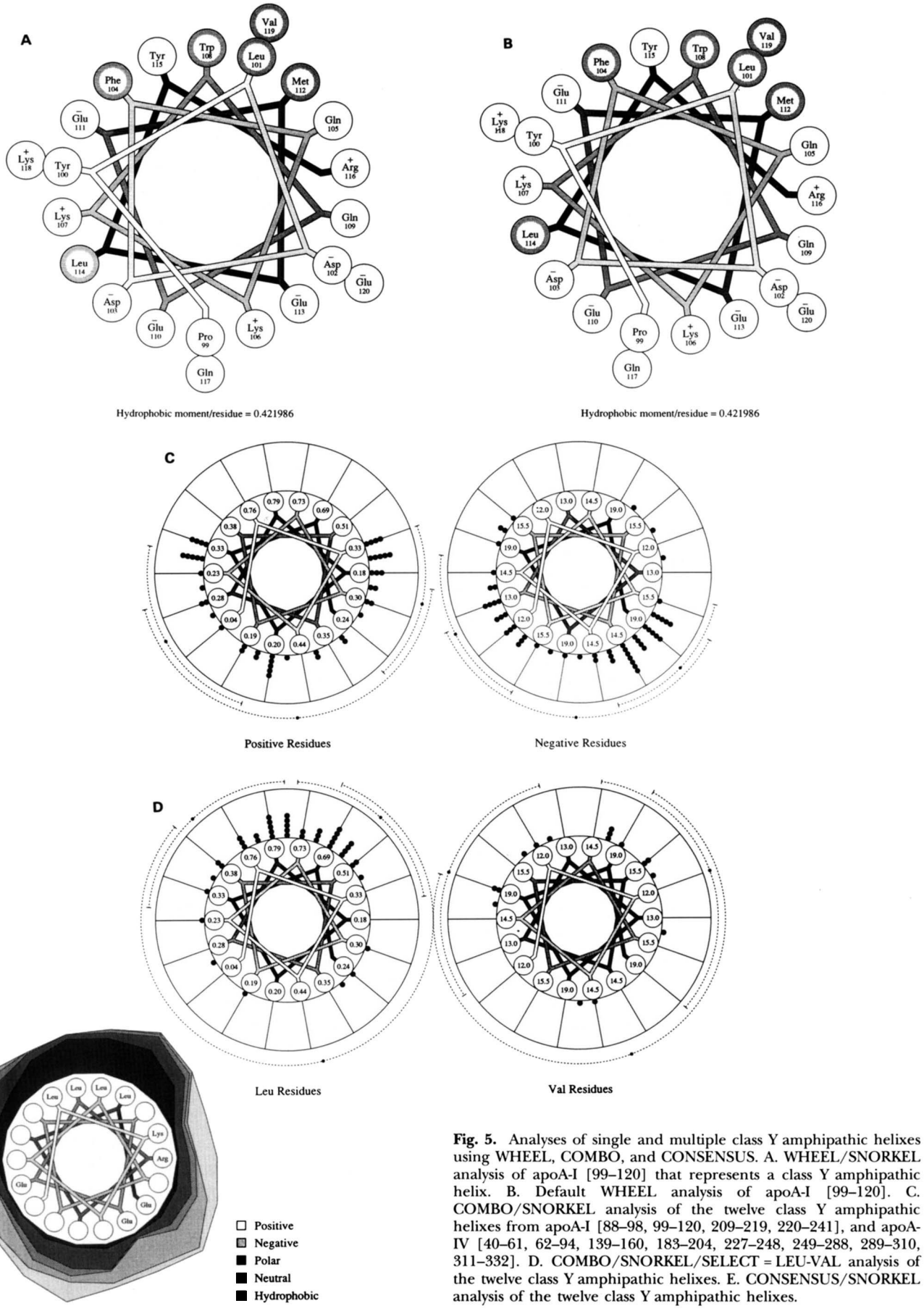

Fig. 5. Analyses of single and multiple class Y amphipathic helixes using WHEEL, COMBO, and CONSENSUS. A. WHEEL/SNORKEL analysis of apoA-I [99–120] that represents a class Y amphipathic helix. B. Default WHEEL analysis of apoA-I [99–120]. C. COMBO/SNORKEL analysis of the twelve class Y amphipathic helixes from apoA-I [88–98, 99–120, 209–219, 220–241], and apoA-IV [40–61, 62–94, 139–160, 183–204, 227–248, 249–288, 289–310, 311–332]. D. COMBO/SNORKEL/SELECT = LEU-VAL analysis of the twelve class Y amphipathic helixes. E. CONSENSUS/SNORKEL analysis of the twelve class Y amphipathic helixes.

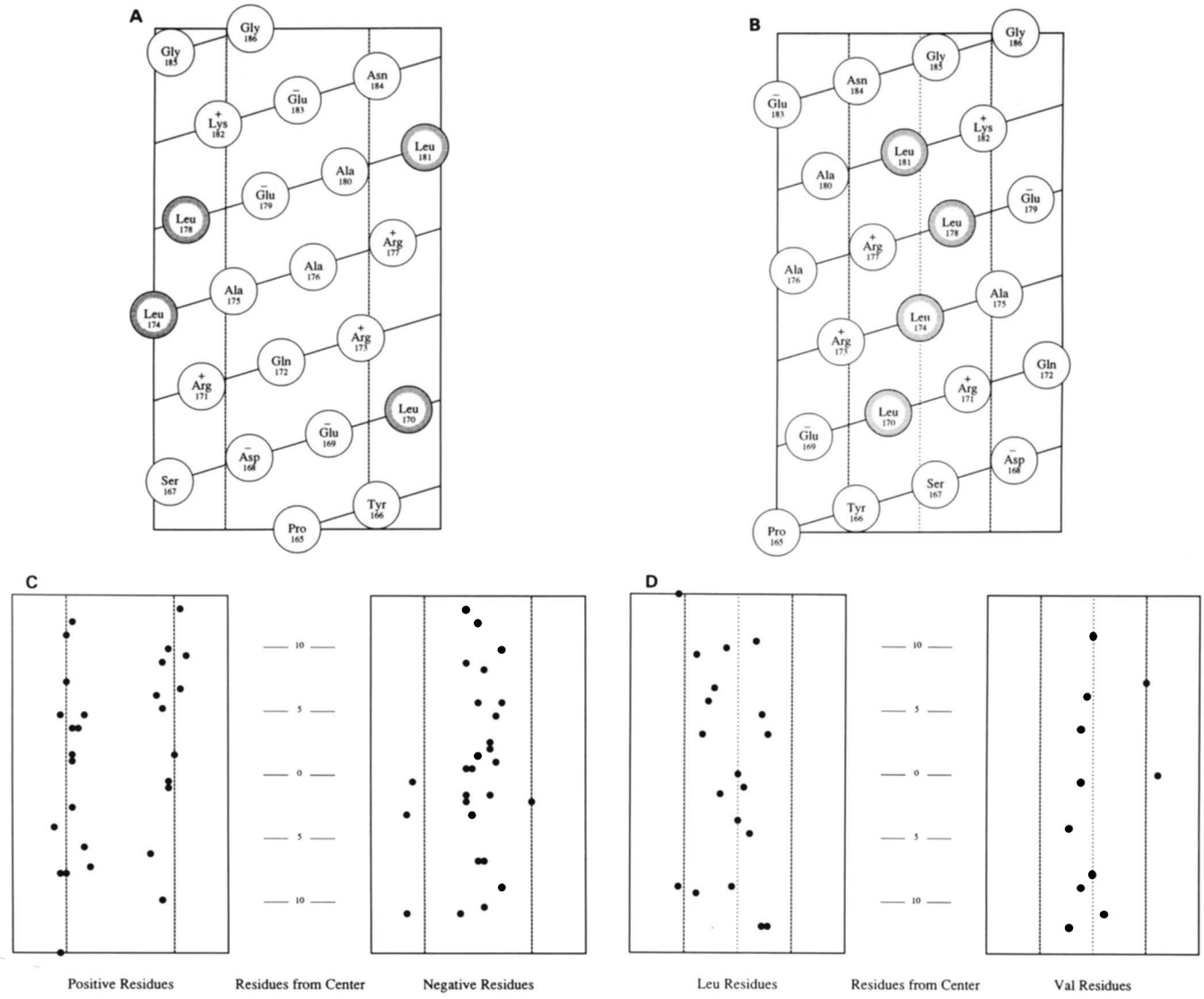

Fig. 6. Analyses of single and multiple class A amphipathic helixes using HELNET and COMNET. A. HELNET/SNORKEL analysis of the class A apoA-I [165–186]. B. Default HELNET analysis of apoA-I [165–186]. C. COMNET/SNORKEL analysis of the seven class A amphipathic helixes listed in Fig. 4. D. COMNET/SNORKEL/SELECT = LEU-VAL analysis of the seven class A amphipathic helixes.

Wheel angles are defined relative to the center of the nonpolar face as + or – 0° to 180°. In this convention the center of the nonpolar face is 0°; + 90° and – 90° are the right and left polar–nonpolar interfaces, respectively. The program quantifies clustering in each of two ways: as one cluster located by the mean of the angles of all the residues (expressed as 0° to 360°) on the wheel and as two clusters located one on the right by the mean of the residue angles between 0° + 180° and one on the left by the mean of the residue angles between 0° and – 180°. For each cluster the mean radial angle and standard deviation (expressed as ± degrees of arc) is calculated and plotted on the positive and negative combo wheel diagrams; the tighter the cluster, the smaller the standard deviation.

Six model amphipathic sequences were generated to serve as cluster standards and the results of COMBO analyses of the sequences are shown in **Fig. 8**. The radial angle (θ) and standard deviation of arc subtended for the two cluster analysis (SD_2) and for the one cluster analysis (SD_1) are displayed above each COMBO analysis.

As an example of tight clustering, two evenly distributed 40° clusters of positive residues were located at the polar–nonpolar interfaces (Figs. 8A and B). This model, analyzed as one cluster, gives a standard deviation of ± 96° of arc; analyzed as two clusters this model gives a standard deviation of ± 18° of arc, with cluster mean radial angles located at + and – 90°.

As an example of zero clustering, positive residues were evenly distributed over 180° of the polar face (Fig. 8F). This model, analyzed as one cluster, gives a standard deviation of ± 59° of arc; analyzed as two clusters this model gives a standard deviation of ± 30°

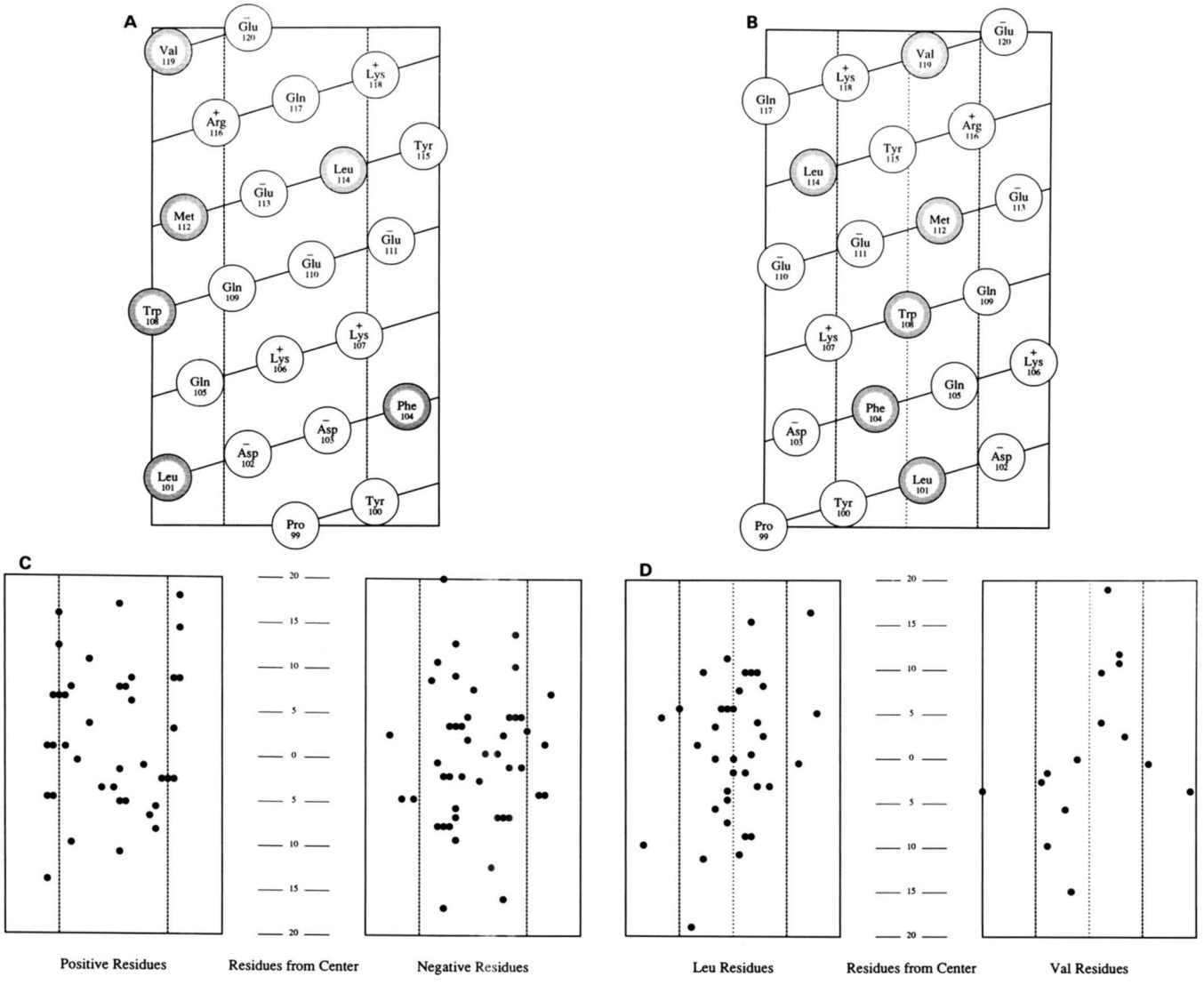

Fig. 7. Analyses of single and multiple class Y amphipathic helixes using HELNET and COMNET. A. HELNET/SNORKEL analysis of the class Y apoA-I [99–120]. B. Default HELNET analysis of apoA-I [99–120]. C. COMNET/SNORKEL analysis of the twelve class Y amphipathic helixes listed in Fig. 5. D. COMNET/SNORKEL/SELECT = LEU-VAL analysis of the twelve class Y amphipathic helixes.

of arc with cluster mean radial angles located at + and – 130°.

Conclusions

We have found the five programs described here, WHEEL, HELNET, COMBO, COMNET, and CONSENSUS, to be helpful tools for analysis of potential amphipathic helixes in primary amino acid sequence data. An accompanying review article (3) bears witness to the usefulness of these programs.

Our future goals are twofold. First, to complete the analyses of the six other classes of amphipathic helixes (class H, polypeptide hormones; class L, lytic polypeptides; class G, globular proteins; class K, calmodulin-regulated protein kinases; class C, coiled-coils, and class M, transmembrane proteins) we described in our previous review article (2). Second, we intend to incorporate new algorithms into our existing programs to identify and classify all amphipathic helical domains in the protein sequence database.

All five computer programs are written in C and their outputs are diagrams specified by a Postscript program. The programs run on DEC VAXes and portability is only limited by the command interface and facilities for automatic printing. All five programs, including options for π helix analysis, are available to interested investigators.

This work was supported in part by National Institutes of Health Grants HL 34343 and AI 28928 to JPS.

Manuscript received 7 November 1991.

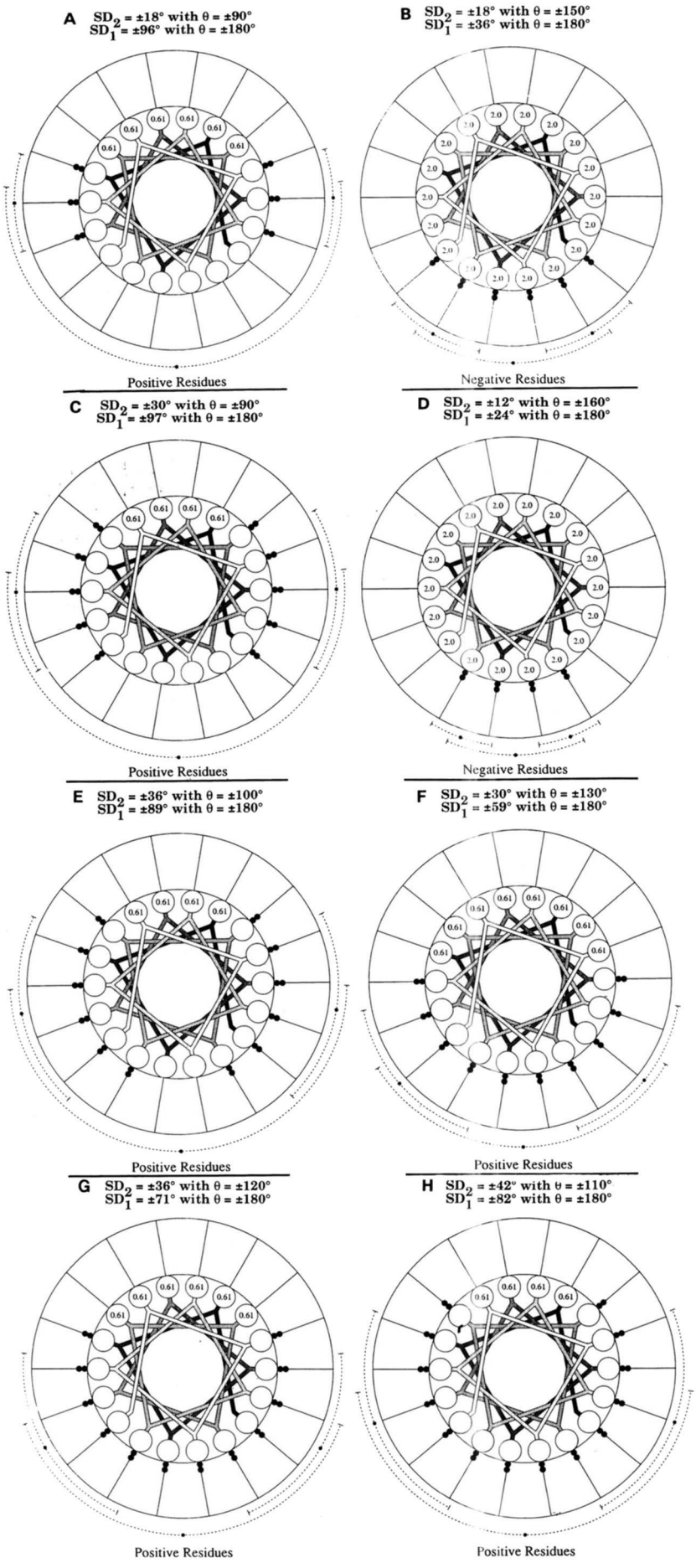

Fig. 8. COMBO analysis of six model amphipathic sequences generated to serve as cluster standards. The radial angle (θ) and standard deviation of arc subtended for the two cluster analysis (SD_2) and for the one cluster analysis (SD_1) are displayed above each COMBO analysis. Each 36 amino acid long sequence is given in single letter code. A and B. Well-defined class A amphipathic helix: EAAEK–AKEKA–EEAKE–KAKEA–AEKAK–EKAEE –AKEKA–K. C and D. Less well-defined class A amphipathic helix: EKKEK–AKEKA–KKAKE–KAKEK–KEKAK–EKAKK–AKEKA–K. E. Very wide positive charge clusters: KKKKK–AKEKA–KKAKE–KAKKK–KKKAK–EKAKK–AKEKA–K. F. Featureless 180° polar face: KAAKK–AKKAA–KKAAK–KAKKA–AKKAK–KAAKK–AAKKA–K. G. Featureless 220° polar face: KAAKK–AKKKA–KKAKK–KAKKA–AKKAK–KKAKK–AKKKA–K. H. Featureless 260° polar face: KKKKK–AKKKA–KKAKK–KAKKK–KKKAK–KKAKK–AKKKA–K.

REFERENCES

1. Segrest, J. P., R. L. Jackson, J. D. Morrisett, and A. M. Gotto, Jr. 1974. A molecular theory of lipid–protein interactions in the plasma lipoproteins. *FEBS Lett.* **38**: 247–253.
2. Segrest, J. P., H. DeLoof, J. G. Dohlman, C. G. Brouillette, and G. M. Anantharamaiah. 1990. Amphipathic helix motif: classes and properties. *Proteins.* **8**: 103–117.
3. Segrest, J. P., M. K. Jones, H. DeLoof, C. G. Brouillette, Y. V. Venkatachalapathi, and G. M. Anantharamaiah. 1992. The amphipathic helix in the exchangeable apolipoproteins: a review of secondary structure and function. *J. Lipid Res.* **33**: 000–000.
4. Schiffer, J., and A. B. Edmundson. 1967. Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. *Biophys. J.* **7**: 121–135.
5. Eisenberg, D., R. M. Weiss, and T. C. Terwilliger. 1982. The helical hydrophobic moment: a measure of the amphipathicity of a helix. *Nature.* **299**: 371–374.
6. Engelman, D. M., T. A. Steitz, and A. Goldman. 1985. Identifying transbilayer helices in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biophys. Chem.* **15**: 321–353.
7. Anantharamaiah, G. M., J. L. Jones, C. G. Brouillette, C. F. Schmidt, B. H. Chung, T. A. Hughes, A. S. Bhown, and J. P. Segrest. 1985. Studies of synthetic peptide analogs of the amphipathic helix. *J. Biol. Chem.* **260**: 10248–10255.
8. Anantharamaiah, G. M. 1986. Synthetic peptide analogs of apolipoproteins. *Methods Enzymol.* **128**: 626–668.
9. Chung, B. H., G. M. Anantharamaiah, C. G. Brouillette, T. Nishida, and J. P. Segrest. 1985. Studies of synthetic peptide analogs of the amphipathic helix: correlation of structure with function. *J. Biol. Chem.* **260**: 10256–10262.
10. Epand, R. M., A. Gawish, M. Iqbal, K. B. Gupta, C. H. Chen, J. P. Segrest, and G. M. Anantharamaiah. 1987. Studies of synthetic peptide analogs of the amphipathic helix. *J. Biol. Chem.* **262**: 9389–9396.
11. Venkatachalapathi, Y. V., K. B. Gupta, H. DeLoof, J. P. Segrest, and G. M. Anantharamaiah. 1990. Positively charged residues, because of their amphipathic nature, can increase the lipid affinity of the amphipathic helix. *In* Peptides: Chemistry and Biology. J. Rivier, editor. ESCOM Press, Leiden. 672–673.
12. Segrest, J. P., Y. V. Venkatachalapathi, S. K. Srinivas, K. B. Gupta, H. DeLoof, and G. Anantharamaiah. 1991. Role of basic amino acid residues in the amphipathic helix: the Snorkel Hypothesis. *In* Ramachandran Festschrift Volume. In press.
13. Lim, V. I. 1978. Polypeptide chain folding through a highly helical intermediate as a general principle of globular protein structure formation. *FEBS Lett.* **89**: 10–14.