A mathematically defined motif for the radial distribution of charged residues on apolipoprotein amphipathic α helixes

Jane B. Hazelrig,* Martin K. Jones,‡ and Jere P. Segrest‡

*Department of Biostatistics, University of Alabama at Birmingham Academic Health Sciences Center; and ‡Atherosclerosis Research Unit, Departments of Medicine and Biochemistry, University of Alabama at Birmingham Medical Center, Birmingham, Alabama 35294 USA

ABSTRACT Multiple amphipathic α-helical candidate domains have been identified in exchangeable apolipoproteins by sequence analysis and indirect experimental evidence. The distribution of charged residues can differ within and between these apolipoproteins. Segrest et al. (Segrest, J. P., H. DeLoof, J. G. Dohlman, C. G. Brouillette, and G. M. Anantharamaiah. 1990. *Proteins.* 8:103–117.) argued that these differences are correlated with lipid affinity. A mathematically defined motif for the particular charge distribution associated with high lipid affinity (class A) is proposed. Primary sequence data from protein segments proposed previously to have an amphipathic α-helical structure are scanned. Counting formulas are presented for determining the conditional probability that the match between an observed charge distribution and the proposed motif would occur by chance. Because the preselected helical segments are short (the modal length is 22) and the motif definition imposes multiple constraints on the acceptable distributions, the computer-based algorithm is quite feasible computationally. 19 of the 20 segments previously assigned to class A match the motif sufficiently well (the remaining one is borderline), while very few others "erroneously" pass the screening test. These results confirm the original assignments of the candidate domains and, thus, support the hypothesis that there is a distinguishable subset of helixes having high lipid affinity. This counting approach is applicable to a growing subset of protein sequence analysis problems in which the segment lengths are short and the motif is complex.

INTRODUCTION

The amphipathic α helix is an important structural motif of proteins and peptides. It is characterized by periodic variations in the hydrophobicity of its residues. If the helix is pictured as a right cylinder, then polar and nonpolar residues are distributed along the long axis on opposite sides of the cylindrical surface.

Apolipoproteins are the protein components of plasma lipoproteins. It was proposed by Segrest et al. (9) that an amphipathic α-helical structure facilitates protein–lipid interactions, and numerous experiments now indicate that it is responsible for the lipid-associating properties of the exchangeable apolipoproteins. Hence, the prediction of this structural motif from primary sequence data could help elucidate the biological function of a newly sequenced protein.

Various algorithms have been proposed for identifying α-helical domains from primary sequence data (2, 3, 6–8). Although their accuracy has not exceeded 70% in the past, there is reason to believe that lipid-associating α-helical domains can be identified more reliably (11). Our database now contains >40 candidate domains. These domains were identified by scanning the primary sequences of the exchangeable apolipoproteins using an empirical algorithm based on visual inspection of helical wheel diagrams (5). The radial arrangement of charged residues about the helical surface appears to differ some between and within apolipoproteins. Segrest et al. (10) presented evidence that the observed differences are correlated with biological attributes such as lipid affinity. Thus it is of interest to use primary sequence data to delineate classes of charge distributions and to determine the class to which a given helix belongs.

This paper deals with the formulation of a statistical model that quantifies the empirical algorithm proposed previously (from visual inspection) for identifying exchangeable apolipoprotein helixes exhibiting high lipid affinity. The question is whether the match between an observed distribution of charges and the model motif is statistically significant. The problem is addressed by counting the total number of possible arrangements (permutations) of positively and negatively charged residues that would satisfy the constraints of the proposed model. In this way it is possible to determine the conditional probability of observing the match by chance.

The model is applied to 100 protein segments, in part, to establish the validity of the original assignment of these candidate domains to a distinct class. The modal length of the sequences is 22, with ~30% of the residues being charged. Thus the computational requirements are well within the power of modern computers. The appropriate counting formulas are presented below together with the statistical evaluations based on their application.

This paper discusses the applicability of this type of analysis to a growing subset of protein sequence analyses in which the motif is complex and the segments are relatively short.

MATHEMATICAL MODEL

Description

This study focuses on those protein segments with secondary structures designated by Segrest et al. (11) as an

Address for correspondence: Jane B. Hazelrig, Ph.D., Department of Biostatistics, BB 145E, UAB Academic Health Sciences Center, Birmingham, AL 35294.

amphipathic α helix. (This designation was obtained by applying the algorithm of Jones et al. [5] to the primary sequence data.) The charged residues of each domain so identified were plotted on helical wheel diagrams in order to visualize the distributions. Graphically, the z-axis was aligned with the long axis of the helix and the amino acid residues were projected onto a circle in the x–y plane. In an idealized α helix, consecutive residues are spaced at 100° intervals. Given a sequence of ≥ 18 residues, there will be at least one residue at each 20° interval. The positive y-axis was assigned 0°. Moving in the clockwise direction, the angles ranged from 0 to 180°. Moving counterclockwise, the angles ranged from 0 to −180°. A hydrophobicity value was assigned to each residue using the GES scale (4). Then the helix was rotated about the z-axis in 20° increments to align the projection of its hydrophobic moment in the x–y plane with 0°. Now the x-axis defines the polar–nonpolar interface of the helix with the polar face lying below the axis.

According to the snorkel hypothesis (10), when the helix is associated with phospholipid, the interfacial basic residues extend ("snorkel") toward the polar face to insert their positively charged moieties into the aqueous milieu. The net effect of the presence of such residues should be an increased lipid affinity. Snorkeling should be facilitated if these interfacial residues are distributed symmetrically about 0°. When the sequence contained at least two positively charged residues, the wheel was rotated again (in 20° increments) if necessary to optimize this symmetry. Note that the first residue in the sequence may be at any of 18 positions after the rotation(s).

Proposed class A motif

The most distinctive feature of amphipathic α helixes in exchangeable apolipoproteins is the clustering of positively charged residues at the polar–nonpolar interface and of negatively charged residues on the polar face, as shown in Fig. 1. Helixes exhibiting the typical charge pattern are hereafter designated as class A. The motif proposed in this paper delineates the charge pattern by specifying that positively charged residues are located at 90 ± 30 and -90 ± 30°, and negatively charged residues are located between either 90 and 180 or −90 and 180°. To satisfy the constraints of the class A motif, only positively charged or uncharged residues should be located at ±60 and ±80°, hereafter referred to as positive residue positions. Only negatively charged or uncharged residues should be located at ±140, ±160, and 180°, hereafter referred to as negative residue positions. Any residue type can be located at ±100 or ±120°, hereafter referred to as free residue positions. No charged residues should be located at 0, ±20, or ±40°, hereafter referred to as uncharged residue positions.

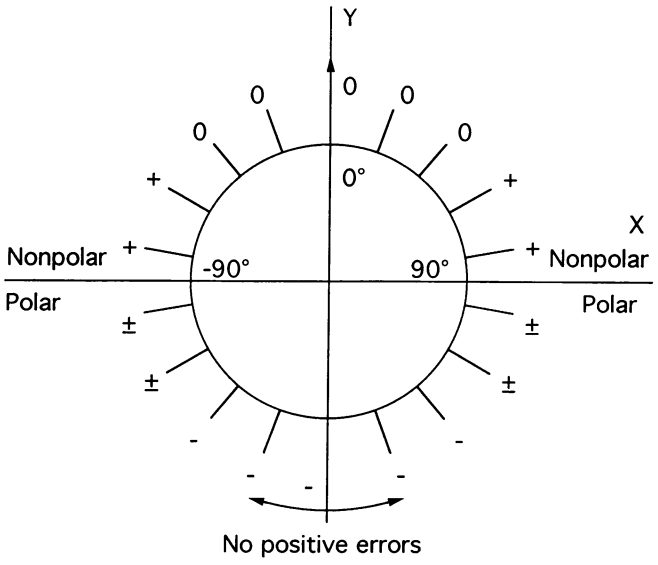

FIGURE 1 Motif for class A exchangeable apolipoprotein amphipathic α helix. Uncharged residues are allowed everywhere. 0, no charged residue allowed; +, positively charged residues allowed; −, negatively charged residues allowed; ±, all residues allowed. Errors are permitted when testing the helix, but no positively charged residues can be located in [−160°, 160°].

Formulation of class membership test

Let

N = number of residues = number of residue positions in sequence
m_p = number of positive residue positions in sequence (nonpolar side of interface, i.e., ±60 and ±80°)
m_n = number of negative residue positions (middle portion of polar face, i.e., ±140, ±160, and 180°)
m_{pn} = number of free residue positions (polar side of interface, i.e., ±100 or ±120°)
m_0 = number of uncharged residue positions (middle portion of nonpolar face, i.e., 0, ±20, or ±40°)
n_p = number of positively charged residues in sequence
n_n = number of negatively charged residues in sequence
n_0 = number of uncharged residues in sequence

By definition

$$N = m_p + m_n + m_{pn} + m_0 = n_p + n_n + n_0.$$

The n's can be determined simply by knowing which residues comprise the candidate helix. Before the m's can be counted, the helix must be aligned, with 0° being assigned to the residue in the central portion of the hydrophobic face. Theoretically this residue should be uncharged. That was the case for all of the helixes reported here.

Suppose a sequence satisfying all the constraints of the class A motif is found. How often would this be observed by chance? Assuming the likelihood that a charged residue will be placed in a positive, negative, free, or uncharged residue position depends only on the number of

residue positions of each type available. Once an uncharged residue is assigned the 0° position, there are $(N-1)$ positions, (m_0-1) uncharged residue positions, and (n_0-1) uncharged residues remaining. Then the question can be addressed as follows. First, count the total number of ways n_p-positive and n_n-negative residues can be distributed among $(N-1)$ residue positions.

$$\binom{N-1}{n_p,\ n_n,\ n_0-1} = \frac{(N-1)!}{n_p!n_n!(n_0-1)!}. \qquad (1)$$

In a hypothetical example where $N = 22$, $n_p = 4$, $n_n = 6$, and $n_0 = 12$, the answer is 74,070,360. Next, count the total number of ways to place n_p positively charged residues in $m_p + m_{pn}$ positions and n_n negatively charged residues in $m_n + m_{np}$ positions.

$$\sum_{i=\max(0,n_p-m_{pn})}^{\min(n_p,m_p)} \binom{m_p}{i} \times \sum_{j=\max(0,n_n-m_{pn})}^{\min(n_n,m_n)} \binom{m_n}{j}\binom{m_{pn}}{n_p-i}\binom{m_{pn}-n_p+i}{n_n-j},$$

where

$$(\mathbf{i}+\mathbf{j}) \geq (\mathbf{n_p}+\mathbf{n_n}-\mathbf{m_{pn}}). \qquad (2)$$

The index i tracks the number of positively charged residues on the nonpolar side of the interface. The maximum possible is the smaller of n_p, the number of positively charged residues, and m_p, the number of positive residue positions. The index j tracks the number of negatively charged residues on the polar face and is similarly limited. Any remaining charged residues must be placed on the polar side of the interface to obey the motif constraints. If there are too many charged residues remaining, these will not fit into the m_{pn} positions available. Therefore i and j must be large enough to ensure a fit.

Suppose in the hypothetical example, $m_p = 5$, $m_n = 6$, $m_{pn} = 6$, and $m_0 = 5$. Then Eq. 2 yields

$$\sum_{i=0}^{4}\binom{5}{i}\sum_{j=0}^{6}\binom{6}{j}\binom{6}{4-i}\binom{2+i}{6-j},$$

with $(i+j) \geq 4$. The sum is 72,660. This is only 0.0981% of the possible number of distributions, so it is not likely to be observed by chance.

Since factors unrelated to lipid affinity could perturb the charge distribution, sequences with a mismatch (say a charged residue located in an uncharged position) will still be class A candidates. Let

$e_p \leq n_p$ = number of positively charged residues on the middle portion of the nonpolar face, and

$e_n \leq n_n$ = number of negatively charged residues on the middle portion of the nonpolarface.

The helix must be rotated before the e's can be counted. Then the number of ways in which no more than e_p and e_n errors can occur is

$$\sum_{k=0}^{e_p}\binom{m_0-1}{k}\sum_{l=0}^{e_n}\binom{m_0-1-k}{l}\sum_{i=\max(0,n_p-k-m_{pn})}^{\min(n_p-k,m_p)}\binom{m_p}{i} \times \sum_{j=\max(0,n_n-l-m_{pn})}^{\min(n_n-l,m_n)}\binom{m_n}{j}\binom{m_{pn}}{n_p-i-k}\binom{m_{pn}-n_p+i+k}{n_n-j-l},$$

where

$$(k+l+i+j) \geq (n_p+n_n-m_{pn}). \qquad (3)$$

Note that the number observed for e_p or e_n could not exceed $m_0 - 1$. Suppose one negative residue is placed on the nonpolar face, that is, $e_n = 1$ and $e_p = 0$. In the hypothetical example, Eq. 3 yields 631,740. Although this is an order of magnitude larger than before, it is still only 0.853% of the total number of ways in which the charged residues could be distributed. Thus, the distribution is unlikely to be observed by chance.

Mismatches could occur in two other ways. A negatively charged residue could be placed on the nonpolar side of the interface, or a positively charged residue could be placed on the polar side of the interface. All these possibilities can be counted, as will be shown below by redefining e_p and e_n as

e_p = number of misplaced positively charged residues

e_n = number of misplaced negatively charged residues.

Before presenting the formula, it is necessary to consider an additional problem. Class Y amphipathic α helixes are found in exchangeable apolipoproteins (11). Class Y helixes also have a cluster of positively charged residues at the polar–nonpolar interface but have, in addition, a third cluster centered on the polar face. The arrangement of these three clusters resembles the letter "Y." Members of this class appear to have some lipid affinity. It is the third cluster of positively charged residues in the center of the polar face that distinguishes class Y from class A. It is difficult to discriminate between these two classes if all types of errors are allowed in the class A model. Therefore, it is desirable to exclude from class A any sequence having one or more positively charged residues at the center of the polar face (±160°, or 180°). This can be accomplished by first excluding from the random model all arrangements that include one or more positive charges in any of these three positions. Let

m_c = number of residue positions at center of polar face.

Then the total number of ways that n_p-positive residues can be distributed among $N - 1 - m_c$ residue positions and n_n-negative residues can be distributed among $N-1$ residue positions is

$$\binom{N-1}{n_p, n_n, n_0-1} - \sum_{i=1}^{\min(m_c,n_p)} \sum_{j=0}^{\min(m_c-i,n_n)} \binom{m_c}{i}\binom{m_c-i}{j} \times \binom{N-1-m_c}{n_p-i}\binom{N-1-m_c-n_p+i}{n_n-j}. \quad (4)$$

Counting all allowed errors, the number of ways in which no more than e_p positive and e_n negative errors can occur is

$$\sum_{q=0}^{\min(e_p,m_n-m_c)} \binom{m_n-m_c}{q} \sum_{r=0}^{\min(e_n,m_p)} \binom{m_p}{r} \times \sum_{k=0}^{\min(e_p-q,m_0-1)} \binom{m_0-1}{k} \sum_{l=0}^{\min(e_n-r,m_0-1-k)} \binom{m_0-1-k}{l} \times \sum_{i=\max(0,n_p-q-k-m_{pn})}^{\min(n_p-q-k,m_p-r)} \binom{m_p-r}{i} \sum_{j=\max(0,n_n-r-l-m_{pn})}^{\min(n_n-r-l,m_n-q)} \binom{m_n-q}{j} \times \binom{m_{pn}}{n_p-i-k-q}\binom{m_{pn}-n_p+i+k+q}{n_n-j-l-r},$$

where

$$(q+r+k+l+i+j) \geq (n_p+n_n-m_{pn}). \quad (5)$$

The ratio of the value calculated from Eq. 5 to that from Eq. 4 expresses the fraction of the total number of allowed arrangements for which no more than e_p-positive and e_n-negative errors would occur after rotation of the helix to align 0° with the hydrophobic moment. Suppose the fraction is 0.05. Then the conditional probability of observing this arrangement (or one with fewer errors) if the candidate is not a member of class A is only 0.05.

Minimum number of charged residues

One of the distinguishing features of amphipathic α helixes in the exchangeable apolipoproteins is their relatively high charge density. At the other end of the spectrum, the primary sequences of some transmembrane protein helical domains contain no charged residues (1, 10). The sum in both Eqs. 4 and 5 is 1 in this case, so that none of these sequences or any of their permutations could be assigned to class A. When is it reasonable to test a sequence? Once the number of each type of residue position has been calculated for an observed sequence, it is possible to determine how many positively and/or negatively charged residues must be present in order for a perfect match not to be attributed to chance. No sequence was tested unless it contained a number of charged residues sufficient to meet this criterion. The sequences in our database of lipid-associating amphipathic helixes range in length from 11 to 64 residues. Calculations show that each must contain four or five charged residues in order for the conditional probability of a chance match to be <0.05, for example. This is not a serious problem when screening for class A helixes since sequences with fewer charged residues can usually be eliminated from class A on the basis of their low charge density.

TABLE 1 Significance of match between class A motif and observed distribution of charged residues

N	n_p	n_n	e_p	e_n	Conditional probability of chance occurrence
26	6	5	0	1	0.00456
21	5	4	1	0	0.0126
26	3	3	0	0	0.0167
13	3	3	0	1	0.0128
28	4	5	0	0	0.00267
24	3	4	0	1	0.0550
12	3	2	0	0	0.0472
21	2	2	1	0	0.205
22	3	3	0	0	0.0224
22	2	6	0	2	0.102
22	4	4	2	0	0.0517
22	4	4	0	0	0.00340
22	4	4	0	0	0.00502
22	4	3	0	0	0.0139
22	4	3	1	0	0.0475
64	13	13	0	0	$0.179 \cdot 10^{-8}$
22	3	4	0	0	0.0139
22	3	2	1	0	0.176
21	4	5	2	1	0.147
22	2	3	0	0	0.0417

Sequences had been assigned previously to class A using helical wheel diagrams. N, residues; n_p (n_n), positively (negatively) charged residues; e_p (e_n), positively (negatively) charged residues not matching class A motif; probability is conditional upon prior rotation of the helix.

RESULTS

The conditional probability that a random sequence having no more than the observed number of class A mismatches would occur by chance is calculated using Eqs. 4 and 5. If there are no positively charged residues centered on the polar face and the conditional probability of chance occurrence is less than a predetermined fraction, the sequence is assigned to class A. Since this is intended to be a screening test for class A helixes, the cutoff is set conservatively at 0.2. All of the sequences tested had been assigned earlier to one of eight classes (11). It should be emphasized that these assignments were made on the basis of indirect experimental evidence (functional and biological attributes) and by visualizing helical wheel diagrams of the primary sequences. 19 of the 20 sequences that were designated previously as class A are assigned to class A by the algorithm defined here as shown in Table 1. The remaining one is barely above the cutoff at 0.205.

There are 11 testable sequences from exchangeable lipoproteins judged earlier to belong to class Y. As shown in Table 2, nine of the sequences are rejected because there are positively charged residues at the center of the

TABLE 2 **Significance of match between class A motif and observed distribution of charged residues**

N	n_p	n_n	e_p	e_n	Conditional probability of chance occurrence
22	3	3	1	0	Reject
11	3	3	1	1	Reject
22	4	6	1	1	Reject
22	4	3	2	1	Reject
22	4	5	1	0	Reject
22	5	4	1	0	Reject
40	6	8	2	0	Reject
22	3	3	1	0	Reject
22	3	5	1	2	0.369
22	3	3	1	0	0.0773
33	7	9	3	2	Reject

Sequences had been assigned previously to class Y using helical wheel diagrams. See Table 1 for definition of column headings. "Reject" indicates positively charged residue(s) centered on the polar face.

polar face, one is rejected because the conditional probability of chance occurrence is >0.2, and one is accepted as class A.

There are eight testable sequences from exchangeable apolipoproteins judged previously to belong to class G*. The results, shown in Table 2, indicate that five of the sequences would be rejected and three would be accepted as class A.

There were two additional types of proteins with lipid-associating amphipathic α helixes described by Segrest et al. (10). Although the sequences are not from exchangeable apolipoproteins, these are compared with the class A motif to test the power of the model to reject a sequence that is not type A. Class H consists of polypeptide hormones. 4 of the 12 sequences are testable, and none is assigned to class A. Class L consists of lytic polypeptides. 5 of the 13 sequences are testable, and none of the 5 is assigned to class A. The results are presented in Table 4.

Amphipathic α helixes are thought to facilitate protein–protein interactions also. Segrest et al. (10) described three protein types. Although these helixes are not associated with lipids, the sequences are compared with the class A motif as a further test of the model. 11 of the 21 sequences in globular α-helical proteins (class G) are testable, and only 1 of these is assigned to class A. Five of the six calmodulin-binding amphipathic helical domains (class K) are testable, and none is assigned to class A. All of the coiled-coil (class C) proteins are testable, and none is assigned to class A. These results are summarized in Table 4. Only 1 of the 33 testable protein sequences thought to have α-helical structures is "incorrectly" identified as belonging to class A.

TABLE 3 **Significance of match between class A motif and observed distribution of charged residues**

N	n_p	n_n	e_p	e_n	Conditional probability of chance occurrence
26	4	5	1	0	Reject
22	3	1	3	0	Reject
26	9	3	4	1	Reject
27	3	5	0	0	0.00404
18	2	3	2	0	0.134
32	4	7	2	4	Reject
26	4	4	3	1	Reject
25	2	3	2	0	0.115

Sequences had been assigned previously to class G* using helical wheel diagrams. See Table 1 for definition of column headings. "Reject" indicates positively charged residue(s) centered on the polar face.

TABLE 4 **Class A motif applied to amphipathic helixes found outside exchangeable apolipoproteins**

Protein type	Number of sequences	Number of testable sequences	Not class A*	Class A*
	Lipid–protein interactions			
Polypeptide hormones	12	4	4	0
Lytic polypeptides	13	5	5	0
	Protein–protein interactions			
Globular α helical proteins	21	11	10	1
Protein kinases	6	5	5	0
Coiled-coil	8	8	8	0
	Totals			
	60	33	32	1

* Class A acceptance based on a conditional probability of chance occurrence < 0.2 and no positively charged residue(s) centered on the polar face.

DISCUSSION

The immediate goal of this study is to formulate a model for evaluating the conditional probability with which the class A motif would be observed by chance when scanning the primary sequences of amphipathic α-helical candidate domains in exchangeable apolipoproteins. The model defines mathematically the motif associated with a subset of sequences designated by Segrest et al. (10) as class A on the basis of indirect experimental evidence of functional and biological attributes (such as lipid affinity) and of the radial distribution of charged residues as visualized on helical wheel diagrams. The results confirm the hypothesis that there is a distinguishable subset of sequences. 19 of the 20 sequences originally assigned to class A are accepted as matching the proposed motif, and the remaining candidate is borderline. At the same time, there are very few "false positives." Only 4 of the other 19 testable helixes in the exchangeable lipoproteins are assigned to class A using the algorithm presented here. The two nontestable helixes could be eliminated from class A on the basis of their low

charge density. Only 1 of the 33 testable sequences in other proteins "incorrectly" passed the screening test, and the 27 nontestable sequences could be eliminated from class A on the basis of their low charge density. Thus the model helps to validate the original assignment of these candidate domains to a distinct class. This, in turn, supports the hypothesis that this motif is associated with those amphipathic α helixes in exchangeable apolipoproteins that exhibit high lipid affinity.

The formulation of a mathematical model offers the usual advantages. Although visual inspection is highly efficient when exploring data, selection based on visualization alone is subjective. By quantitatively defining a motif for a perceived pattern, it is possible to develop computer-based algorithms to search primary sequence data for matches and to assess objectively the conditional probability that such a match would arise by chance.

This study is of more general interest because the strategy is applicable to a growing subset of the field of protein sequence analysis. A major focus of this field has been on the identification of a statistically significant sequence pattern in a segment or segments of a protein that is several hundred residues long. Statistical evaluations have been based on theoretical models where possible. When a theoretical model appropriate for the pattern and data was not available, investigators have relied on permutation procedures. The observed residues, or some defined subset of these residues, were randomly shuffled and the reconstructed sequence was scanned for the original sequence pattern of interest. Sometimes a compromise has been necessary between the number of permutations that can feasibly be generated and scanned and the number required for the statistical accuracy desired.

The increasing success of these efforts is making it possible for investigators to collect a significant number of protein segments that exhibit a defined pattern. Naturally, the segments are short compared with the length of whole proteins. Although members of this set are distinguishable from the general protein population, often the segments are still heterogeneous with regard to some biological attribute(s). In such cases it is logical to seek subpatterns that may correspond to biological attributes. The motifs defining these subpatterns will be more complex than the one delineating the original pattern. The complexity of the motif coupled with the relative shortness of the segments makes it unlikely that theoretical models will be available for statistical evaluations.

However, it is these very factors that render attractive the approach presented here. Suppose the amino acid residues are grouped into classes according to some attribute, such as charge. It is computationally easy to enumerate the ways that the observed number of residues in each class can be distributed over a sequence of observed length N, using a formula like that in Eq. 1 or 4. Now it is necessary only to count the number of those arrangements that match the motif with no more than the observed number of errors. When a relatively small percentage of the data permutations can meet the complex constraints of the motif, this is quite feasible for short segments (say $\leq$50 residues). Therefore, it is possible to calculate exactly the conditional probability with which an observed sequence meeting the constraints of a proposed model (such as the class A motif) would occur by chance. Similar results could be obtained by permutation analysis. However, for a fixed degree of accuracy, the approach presented here should be more efficient when the sequences are relatively short and the motif is complex. This conjecture can be tested as more data become available.

We acknowledge the helpful comments of the reviewers.

This work was supported in part by National Institutes of Health grants H6-34343 and AI-28928.

Received for publication 9 June 1992 and in final form 8 February 1993.

REFERENCES

1. Allen, J. P., G. Feher, T. O. Yeates, H. Komiya, and D. C. Ross. 1987. Structure of the protein subunits in the photosynthetic reaction center of *Rhodobacter sphaeriodes* R-26: the protein subunits. *Proc. Natl. Acad. Sci. USA.* 84:6162–6166.
2. Boguski, M. S., N. A. Elshourbagy, J. M. Taylor, and J. L. Gordon. 1985. Comparative analysis of the repeated sequences in rat apolipoprotein A-I, A-IV, and E. *Proc. Natl. Acad. Sci. USA.* 82:992–996.
3. Eisenberg, D., R. M. Weiss, and T. C. Terwilliger. 1984. The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc. Natl. Acad. Sci. USA.* 74:2677–2681.
4. Engleman, D. M., T. A. Steitz, and A. Goldman. 1985. Identifying translipid helixes in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biophys. Chem.* 15:321–353.
5. Jones, M. K., G. M. Anantharamaiah, and J. P. Segrest. 1992. Computer algorithms to identify and classify amphipathic α helical domains. *J. Lipid Res.* 33:287–296.
6. Kubota, Y., S. Takahashi, K. Nishikawa, and T. Ooi. 1980. Homology in protein sequences expressed by correlation coefficients. *J. Theor. Biol.* 91:347–361.
7. Lim, V. I. 1974. Polypeptide chain folding through a highly helical intermediate as a general principle of globular protein structure formation. *J. Mol. Biol.* 88:873–894.
8. Schiffer, M., and A. B. Edmundson. 1967. Use of helical wheels to represent the structures of protein and to identify segments with helical potential. *Biophy. J.* 7:121–135.
9. 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 (Fed. Eur. Biochem. Soc.) Lett.* 38:247–253.
10. 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.
11. 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.* 33:141–166.