Biochem. J. (1990) 269, 691–696 (Printed in Great Britain)

The distribution of physical, chemical and conformational properties in signal and nascent peptides

M. PRABHAKARAN*
Department of Medicine and the Atherosclerosis Research Unit, University of Alabama at Birmingham, Birmingham, AL 35294, U.S.A.

Signal peptides play a major role in an as-yet-undefined way in the translocation of proteins across membranes. The sequential arrangement of the chemical, physical and conformational properties of the signal and nascent amino acid sequences of the translocated proteins has been compiled and analysed in the present study. The sequence data of 126 signal peptides of length between 18 and 21 residues form the basis of this study. The statistical distribution of the following properties was studied: hydrophobicity, M_r, bulkiness, chromatographic index and preference for adopting α-helical, β-sheet and turn structures. The contribution of each property to the sequence arrangement was derived. A hydrophobic core sequence was found in all signal peptides investigated. The structural arrangement of the cleavage site was also clearly revealed by this study. Most of the physical properties of the individual sequences correlated (correlation coefficient ~ 0.4) very well with the average distribution. The preferred occupancy of amino acid residues in the signal and nascent sequences was also calculated and correlated with their property distribution. The periodic behaviour of the signal and nascent chains was revealed by calculating their hydrophobic moments for various repetitive conformations. A graphical analysis of average hydrophobic moments versus average hydrophobicity of peptides revealed the transmembrane characteristics of signal peptides and globular characteristics of the nascent peptides.

INTRODUCTION

It has been firmly established that signal peptides are highly essential for protein translocation through the endoplasmic reticulum (Briggs & Gierasch, 1986; Walter & Lingappa, 1986), and many of their sequences have been identified by DNA sequencing methods. Sequence studies have also shown the existence of a hydrophobic core in the signal sequences (von Heijne, 1981, 1985; Pincus & Klausner, 1982). Various theories have been proposed (Blobel & Dobberstein, 1975*a*,*b*; von Heijne & Blomberg, 1979; Wickner, 1981; Inouye & Halegoua, 1980; Nesmeyanova, 1982; Randall & Hardy, 1984) to explain the mechanism of protein translocation and the role played by the signal peptide. In order to understand those features that are important in the translocation process, I have studied the distribution of various properties of amino acid residues along the signal sequences and also along the adjacent portion of the mature chain. Extensive calculations on signal peptides have previously been reported, with emphasis on the hydrophobic nature of the central core. Nevertheless the present study is justified on the following grounds: (1) access to a larger data basis from cDNA studies; (2) a comprehensive view of chain properties of the signal region; (3) a comparison of properties of the signal region with those of the adjacent portions of the mature chain; (4) establishment of rules for the design of a consensus sequence for the signal region.

MATERIALS AND METHODS

The sequence data of 290 eukaryotic proteins that are transported through the endoplasmic reticulum are derived from the translation of DNA data received from the gene bank of the University of Wisconsin after elimination of redundant and closely similar structures to avoid duplication. This study considers only 126 sequences whose lengths are limited within 18 and 21 residues for effective comparison. The property distributions from the cleavage site towards the *N*-terminal of the signal peptide (-1 to -21) from available data and on the nascent chain ($+1$ to $+21$) are summed up for all the sequences and the average distribution of each property and standard deviation are obtained. The standard deviations are plotted as the upper and lower lines along with the mean. The preference factor for acidic and basic amino acid residues is calculated as the ratio of the observed to the expected distribution. The following properties are studied: hydrophobicity, chromatographic index, M_r, bulkiness and relative frequency of occurrence of α-helical, β-sheet and turn structures. The values of the various properties considered in this study are presented in Table 1. In order to compare the relative extents of variations in signal and nascent peptides the coefficient variation (Sokal & Rohlf, 1969), the standard deviation expressed as a percentage of mean, is calculated. The values of the hydrophobic moment of a peptide of N residues in which side chains protrude perpendicular to the axis at regular δ intervals is given by:

$$\mu_H = \left\{\left[\sum_i H_i \cdot \cos(i\delta)\right]^2 + \left[\sum_i H_i \cdot \sin(i\delta)\right]^2\right\}^{\frac{1}{2}}$$

in which H_i is the hydrophobicity of the ith residue and δ is 100° for α-helix and 160–180° for β-sheets.

RESULTS AND DISCUSSION

Distribution of chemical parameters

There is a considerable diversity of opinion concerning the appropriate choice of hydrophobicity scale, especially with regard to transmembrane proteins. Scales have been derived on the basis of solubility measurements, vapour pressure of side-chain analogues and analysis of side-chain distribution in soluble proteins. In the present study Engelman's scale (Engelman *et al.*,

* Present address: Department of Medicinal Chemistry and Pharmacognosy (M/C 781), University of Illinois at Chicago, Box 6998, Chicago, IL 60680, U.S.A.

Table 1. Physical, chemical and conformational properties of amino acids/residues

Parameters: H, hydrophobicity (Engelman scale); M_r, relative molecular mass (weight average); B_1, bulkiness; R, chromatography index; α, relative frequency of occurrence in α-helix structure; β, relative frequency of occurrence in β-sheet structure; Tu, relative frequency of occurrence in reverse turns.

Amino acid	H (kJ/mol)	M_r	B_1 (nm^2)	R	α	β	Tu
Ala	−6.7	89	0.1150	9.9	1.29	0.90	0.78
Asp	38.5	133	0.1168	2.8	1.04	0.72	1.41
Cys	−8.4	121	0.1346	2.8	1.11	0.74	0.80
Glu	34.3	147	0.1357	3.2	1.44	0.75	1.00
Phe	−15.5	165	0.1980	18.8	1.07	1.32	0.58
Gly	−4.2	75	0.0340	5.6	0.56	0.92	1.64
His	12.6	155	0.1367	8.2	1.22	1.08	0.69
Ile	−13.0	131	0.2140	17.1	0.97	1.45	0.51
Lys	36.8	146	0.1571	3.5	1.23	0.77	0.96
Leu	−11.7	131	0.2140	17.6	1.30	1.02	0.59
Met	−14.2	149	0.1625	14.7	1.47	0.97	0.39
Asn	20.1	132	0.1282	5.4	0.90	0.76	1.28
Pro	0.8	115	0.1743	14.8	0.52	0.64	1.91
Gln	17.2	146	0.1445	9.0	1.27	0.80	0.97
Arg	51.5	174	0.1428	4.6	0.96	0.99	0.88
Ser	−2.5	105	0.0947	6.9	0.82	0.95	1.33
Thr	−5.0	119	0.1577	9.5	0.82	1.21	1.03
Val	−10.9	117	0.2157	14.3	0.91	1.49	0.47
Trp	−7.9	204	0.2161	17.0	0.99	1.14	0.75
Tyr	2.9	181	0.1803	15.0	0.72	1.25	1.05

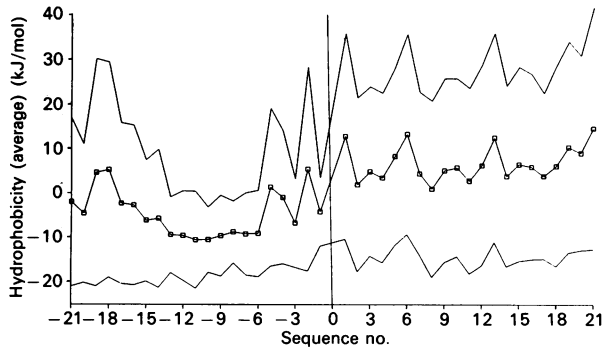

Fig. 1. Average hydrophobic parameter (Engelman scale) distribution

The central line in the figures represents the average, whereas the upper and lower lines show the standard deviation from the average. The central vertical line demarcates signal and nascent peptides.

1986), which describes transmembrane helices, is used. Fig. 1 shows the average hydrophobic distribution. The distribution shows a hydrophobic core over the positions −6 to −13. There is considerably less variation in hydrophobicity over this region than elsewhere, as shown by the relatively small deviations here. The oscillatory behaviour of hydrophobicity near the cleavage site (positions −6 to +1) is clear, as is the polar character of the N-terminal side, except the starting site position 1. Table 2 shows the average coefficient of variation of various properties over the 21 residues in the mature and signal sequences. The near-random distribution of the mature chain side is evident from the large values of coefficient variation of the properties. Even though the polar nature of the N-terminal side is demonstrated statistically, there is considerable breadth to the distribution, as revealed from standard deviation values. Fig. 1 also shows that the hydrophobic core of the signal peptides is not evenly matched by a hydrophilic core of the nascent chain.

Distribution of physical parameters

The distribution of M_r as a function of sequence position is shown in Fig. 2. This plot reveals that positions −4, −3 and −1 are preferred by low-M_r residues, whereas positions −2 and +1 are preferred by residues of high M_r. The favoured occurrence of alanine, glycine and serine at position −1 and disfavoured occurrence of glycine and serine at position +1 are the reasons for this distribution. This unique distribution could be responsible for the recognition of this site by cleavage peptidase. This observation is consistent with the cleavage rule proposed by Pearlman & Halvorson (1983) and von Heijne (1984). The bulkiness of an amino acid is defined as the ratio of the side-chain volume to length, which provides a measure of average cross-section of the amino acid, thus having relevance to packing considerations (Zimmerman *et al.*, 1968). The distribution of bulkiness, as shown in Fig. 3, shows a difference in the distribution pattern between the signal and nascent regions. The bulkiness distribution shows that residues with larger bulkiness are preferred in the hydrophobic core to be accommodated easily in the lipid environment. The −4 and −1 positions are favoured by residues of low bulkiness and low M_r. However, bulkiness increases at positions +1 and +2. At position −3 residues of low M_r are favoured, but not those of low bulkiness. The change in the distribution pattern of M_r is exhibited dramatically at the cleavage-site region, especially in comparison with the distribution of bulkiness. The change in the distribution pattern of both M_r and bulkiness at the cleavage site shows that the size and shape of the residues play a critical role at this site. The chromatographic index (Zimmerman *et al.*, 1968) of an amino

Table 2. Coefficients of variation and correlation coefficients between average and exact distribution of properties in signal and nascent peptides

For definition of parameters see Table 1. Positions of residues are indicated in parentheses.

	Signal peptides				Nascent peptides			
	Coefficient of variation			Correlation	Coefficient of variation			Correlation
Parameter	Average	Maximum	Minimum	coefficient	Average	Maximum	Minimum	coefficient
H	410.9	1527.8 (−4)	70.1 (−10)	0.379	469.5	2090.3 (8)	170.1 (6)	0.184
M_r	19.4	31.7 (−4)	11.0 (−21)	0.383	22.2	32.0 (16)	15.9 (11)	0.198
B_1	34.5	83.8 (−4)	18.0 (−21)	0.441	33.5	53.2 (16)	25.6 (13)	0.277
R	44.7	73.7 (−4)	30.2 (−13)	0.415	59.0	77.1 (1)	48.9 (2)	0.246
α-Helix	24.0	40.9 (−4)	16.7 (−11)	0.327	29.0	41.2 (8)	22.8 (19)	0.230
β-Sheet	23.7	27.9 (−5)	18.7 (−20)	0.323	24.3	29.6 (2)	21.1 (17)	0.228
Turn	49.8	56.5 (−12)	41.0 (−2)	0.391	41.5	51.9 (2)	32.2 (1)	0.229

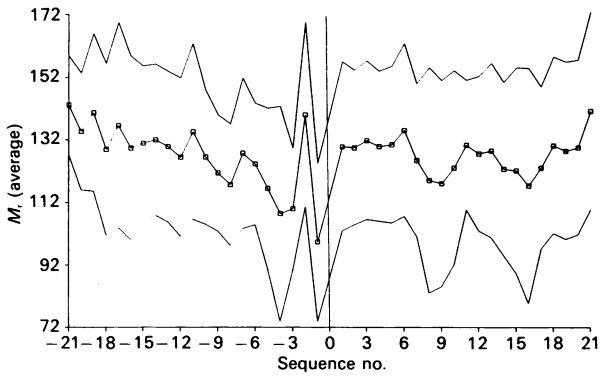

Fig. 2. Distribution of M_r

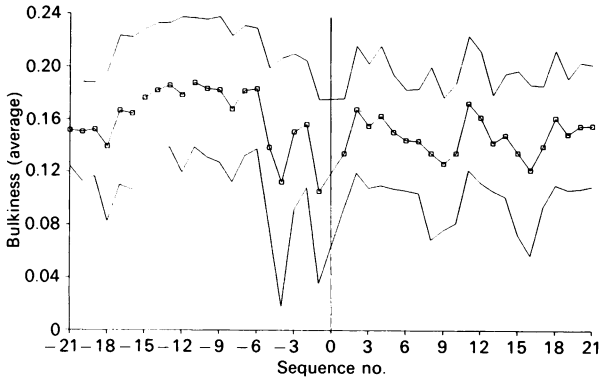

Fig. 3. Distribution of bulkiness

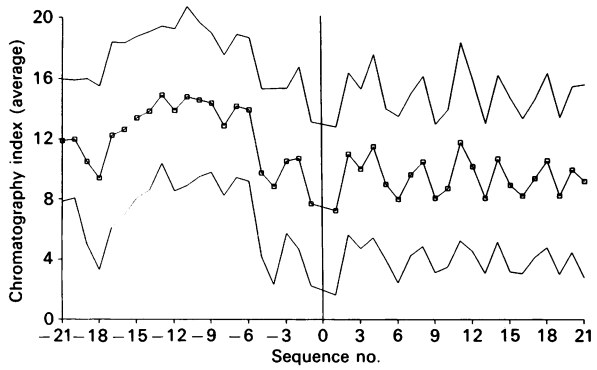

Fig. 4. Distribution of chromatography index

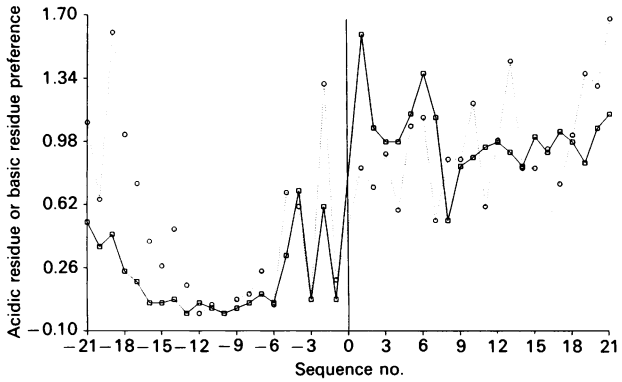

Fig. 5. Distribution of acidic residue (——) and basic residue (·····), preference

acid specifies its characteristic migration rate in a solvent/absorbent system that is a measure of the composite nature of interaction of solute, solvent and hydrophobic absorbent. Both chromatographic index and hydrophobicity could be considered as different hydropathy measures of amino acid residues. Both measures are used here in view of the low correlation (correlation coefficient −0.2) between these two indices. The distribution of the chromatographic index (Fig. 4) reveals the average migration flow of the various segments in a hydrophobic lipid environment. This plot clearly reveals that the nascent chains need external assistance to migrate through membrane. M_r, bulkiness and chromatography index exhibit similar magnitudes in the coefficient of variations in both signal and nascent peptides. The favoured occurrence of methionine at position −21 leads to the minimum variation at that site for these parameters. It may also be noted that the maximum occurs at position −4 for the coefficient of variation. For these three properties the standard deviations also portray similar trends in both nascent and signal peptide sites.

Distribution of charged residues

The numbers of acidic (aspartic acid and glutamic acid) as well as basic (lysine, arginine and histidine) residues distributed in each sequence position are calculated. These numbers are divided by the average number of acidic and basic residues respectively in the sequence position to derive their preference at those sites. Fig. 5 is the preference plot for acidic as well as basic residues in each sequence position. There is a tendency for basic residues to be in the *N*-terminal region of the signal peptides and at positions +13, +19 and +21. The acidic residues tend to be at the cleavage site. Both types of residue show high preference at the nascent sites in comparison with signal peptides. It should be noted that the preference of acidic residues is always less than 1 in the signal-peptide region. A comparatively high acidic preference at the +1 position and the distribution of basic residues in the *N*-terminal region could play a significant role in translocation as well as cleavage.

Distribution of conformational parameters

The distributions of conformational preferences of amino acid residues in the signal and nascent peptides are shown in Figs. 6–8. These parameters are taken from Creighton (1983). Residues in the continuous stretch from positions −21 to −6 have an average α-helical preference value greater than 1.0. The preference of α-helix at the cleavage site is less than 1. There is no preferential distribution at the nascent site. It may be noted that the β-sheet preference for the same hydrophobic core is always greater than 1, whereas the turn preference is always less than 1. The preference for turn is evident only at the cleavage site. The coefficients of variations are found to be minimum for α-helix and β-sheet structures in the hydrophobic core, whereas the minimum for turns is found at the cleavage site. The similar values of the preference for α-helix and β-sheet structures for the hydrophobic core lead to the conclusion that structural transitions may be possible within these regions of the chain, especially as the chain passes through the lipid environment of the membrane of the endoplasmic reticulum. β-Sheet conformations for signal peptides have also been considered in protein export (Randall & Hardy, 1989). Within a lipid bilayer there may be significant changes in the conformational preference factors compared with aqueous solution, and by using these parameters the secondary structures could be predicted only with a 50% accuracy.

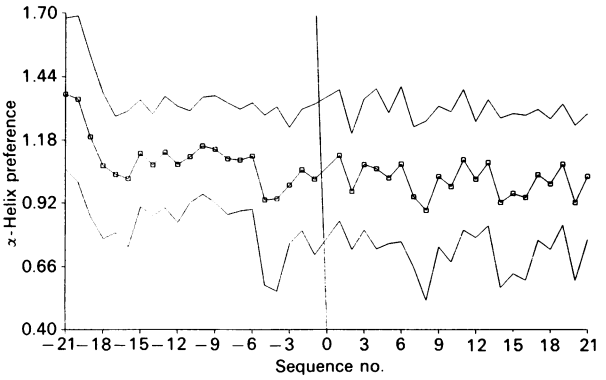

Fig. 6. Distribution of α-helix preference

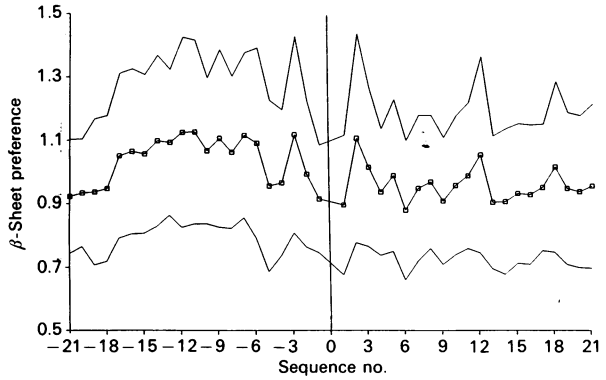

Fig. 7. Distribution of β-sheet preference

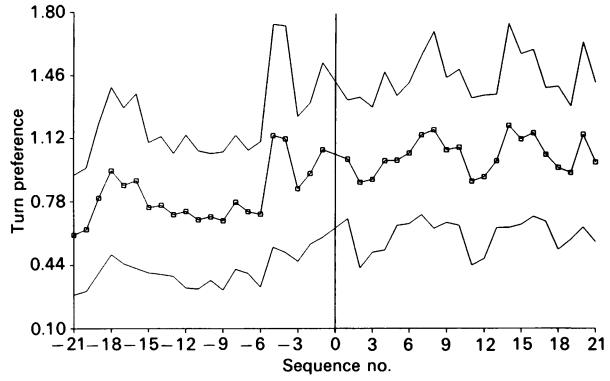

Fig. 8. Distribution of turn preference

PROPERTY CORRELATION

If the above properties have a determining influence on the sequence arrangement and thereby on the function of signal peptides, the average distribution should have a high correlation with the individual distribution. This influence is tested by calculating the correlation between the distribution of average properties of signal peptides and the actual distribution of these properties. The correlations, presented in Table 2, are significant even at the 0.1 level. Bulkiness shows the highest correlation and β-sheet shows the lowest correlation. The correlation values reveal the influence of these parameters in the present study in the choice and the arrangement of amino acid residues in signal peptides. However, no property has any over-riding influence. In particular, the presence of a central hydrophobic core is clearly revealed in these studies. However, the contribution of hydrophobicity to the arrangement of a sequence is on a par with other properties.

Distribution of amino acid residues

The property distribution is due to the difference in the amino acid occupancy level at the various sequence positions in the signal and nascent peptides. Table 3 shows the first four or most probable residues in each position. The preferential factors for the residues are given in parentheses. A perusal of this Table shows the relatively high level of occupancy for leucine from positions −6 to −16. The hydrophobic core is occupied mostly by leucine, valine, phenylalanine and alanine. Since methionine is required as the starting residue, the high occupancy of methionine in the *N*-terminal region is an obvious one. At the *N*-terminal side the basic residues arginine and lysine show high preference, which contributes to α-helix inhibition at these positions and possible electrostatic interactions with the

Table 3. Preferential amino acid residues in signal and nascent peptides

−1 is the starting residue in signal peptides from cleavage site and +1 is the starting residue in nascent chains. The numbers within parentheses denote their preferences with respect to their occupation in globular proteins.

Position	Preference I	II	III	IV
−21	Met (38.46)	Asp (2.44)	Ala (0.99)	Arg (0.86)
−20	Met (30.29)	Ala (1.96)	Gly (0.96)	Leu (0.74)
−19	Met (19.58)	Arg (2.41)	Lys (1.43)	Asn (1.03)
−18	Met (8.58)	Arg (1.92)	Rtp (1.76)	Ala (1.72)
−17	Trp (7.06)	Phe (2.62)	Leu (2.06)	Cys (1.79)
−16	Phe (2.80)	Leu (2.41)	Trp (2.20)	Ala (1.41)
−15	Leu (4.13)	Cys (2.98)	Phe (2.10)	Trp (1.76)
−14	Leu (4.21)	Trp (2.20)	Phe (1.92)	Ile (1.80)
−13	Leu (5.07)	Trp (3.53)	Ile (2.29)	Ala (1.61)
−12	Leu (3.61)	Phe (3.15)	Val (2.61)	Cys (2.09)
−11	Phe (5.60)	Leu (4.21)	Cys (3.58)	Trp (1.32)
−10	Leu (6.02)	Cys (2.98)	Phe (2.62)	Ile (1.14)
−9	Leu (4.81)	Val (2.72)	Ala (2.02)	Phe (1.40)
−8	Leu (4.47)	Cys (2.38)	Ala (2.22)	Val (1.52)
−7	Leu (4.04)	Trp (3.53)	Ile (1.80)	Phe (1.75)
−6	Leu (3.95)	Cys (2.68)	Val (2.39)	Ala (2.02)
−5	Pro (3.01)	Thr (2.74)	Ala (1.82)	Ser (1.45)
−4	Gly (4.69)	Ala (1.92)	Cys (1.49)	Gln (1.40)
−3	Val (3.37)	Ala (3.23)	Cys (2.68)	Thr (2.37)
−2	Trp (3.53)	His (3.26)	Gln (2.45)	Arg (1.75)
−1	Ala (5.15)	Cys (4.77)	Ser (2.80)	Gly (2.20)
1	Gln (4.20)	Cys (2.98)	Glu (2.22)	Asp (1.73)
2	Val (3.26)	Ile (2.78)	Cys (2.09)	Tyr (1.86)
3	Gln (4.37)	Cys (1.79)	His (1.78)	Ile (1.63)
4	Gln (2.97)	Leu (2.84)	Pro (2.56)	Met (2.34)
5	Thr (3.11)	Gln (2.62)	Asn (1.47)	Glu (1.33)
6	Gln (6.47)	Cys (3.87)	Glu (2.22)	Arg (1.92)
7	Ser (3.35)	Pro (2.56)	Tyr (2.32)	Glu (1.45)
8	Gly (2.97)	Pro (2.71)	Phe (1.75)	Leu (1.29)
9	Cys (2.98)	Ala (2.42)	Gly (1.72)	Gln (1.57)
10	His (2.08)	Ser (1.81)	Cys (1.79)	Glu (1.67)
11	Leu (3.18)	Pro (1.81)	Val (1.52)	Gln (1.40)
12	Val (3.15)	Ile (1.63)	Glu (1.33)	Asn (1.31)
13	Gln (2.10)	Lys (1.94)	Ala (1.92)	Arg (1.75)
14	Pro (4.81)	Met (2.73)	Ser (1.54)	Trp (1.32)
15	Gly (2.39)	Leu (1.72)	Cys (1.49)	Arg (1.40)
16	Gly (3.35)	Cys (1.49)	Asn (1.47)	Gln (1.22)
17	Ser (2.62)	Gln (1.75)	Ala (1.51)	Glu (1.33)
18	Val (2.18)	Tyr (1.86)	Leu (1.81)	Arg (1.57)
19	Cys (2.62)	Lys (2.57)	Val (1.77)	Ala (1.65)
20	Pro (3.57)	Arg (1.89)	Lys (1.72)	Ser (1.36)
21	Arg (3.45)	Tyr (2.29)	Met (1.92)	Asp (1.83)

phospholipid head. The acidic residues aspartic acid (at position −21) and glutamic acid (at position −20) are also preferred at the *N*-terminal. At position −5 the residues proline, threonine, alanine and serine are highly preferred, paving the way for a possible bend at that position. At position −1 the low-M_r residues alanine, glycine and serine are preferred. The position +1 is highly occupied by the acidic residues aspartic acid and glutamic acid. The number of cysteine residues may be numerically small, but their frequency of occupation is relatively high at the cleavage site. At the mature site the occupancy of cysteine at positions at +1 and +2, the preference for polar residues at position +1, the non-polar preference at position +2 and the preferential occupation of glutamine at various positions are clearly visible from Table 3. The values of the preferential factors at the mature site are lower than at the signal site. This Table of preferred residues should be highly useful for protein engineers in the design of consensus signal peptides.

Hydrophobic moment and secondary structure

Eisenberg *et al.* (1984*a*), from the hydrophobic moment calculations, suggest that the periodicity of the hydrophobicity of the protein secondary structure is a factor in the formation of secondary structure. The conformational preference of signal peptides is calculated from the distribution pattern of conformational parameters derived mainly from water-soluble globular proteins. Their conformational preference in membrane environment is calculated from the hydrophobic moment value of the signal and nascent peptides of equal size for all the sequences under consideration by varying the periodicity of the residues by changing the successive angle values of the residues in a growing chain. Thus for an α-helix the angle is 100° (3.6 residues per turn) and for a β-sheet structure the angle is expected to be in the range 160–180° (2.3 to 2.1 residues per turn). Figs. 9(*a*) and 9(*b*) display the total hydrophobic moment for signal and nascent peptides respectively for the periodic angle from 80° to 180°. The relative values and type of variations within and between the signal and nascent peptides are considered in this analysis. In the case of nascent peptides, a peak was observed both in the α-helical region as well as in the β-sheet regions, leading to the inference of stability of these globular regions mostly from these two secondary structures. The moment values for the nascent peptides are higher than for the signal peptides. The uniform hydrophobic core has smaller standard deviation in comparison with nascent peptides. The detection of maxima is found for signal peptides at four regions (80°, 110°, 130° and 170°) in comparison with nascent peptides. It is also to be noted that two stable regions (110° and 170°) fall not exactly at the α-helix and β-sheet regions. This leads to the inference that signal peptides can be stabilized hydrophobically in various conformations, and transitions could be possible between these relatively unstable structures. This is also confirmed from the analysis of the secondary-structure preference of the hydrophobic core.

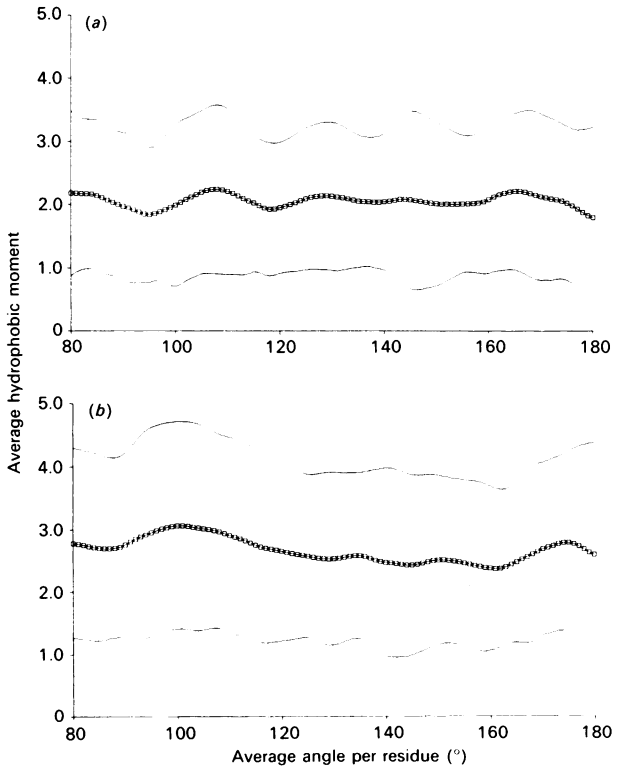

Fig. 9. Average hydrophobic moment for signal peptides (*a*) and nascent peptides (*b*) for different structural conformations

Mean hydrophobic moment and mean hydrophobicity

The preference for hydrophobic environment of the signal peptides is clearly evident from this study. The next question that should be addressed is the conformational arrangement of these signal and nascent peptides in a membrane environment. The calculation of mean hydrophobic moment and mean hydrophobicity by using the algorithm of Eisenberg *et al.*

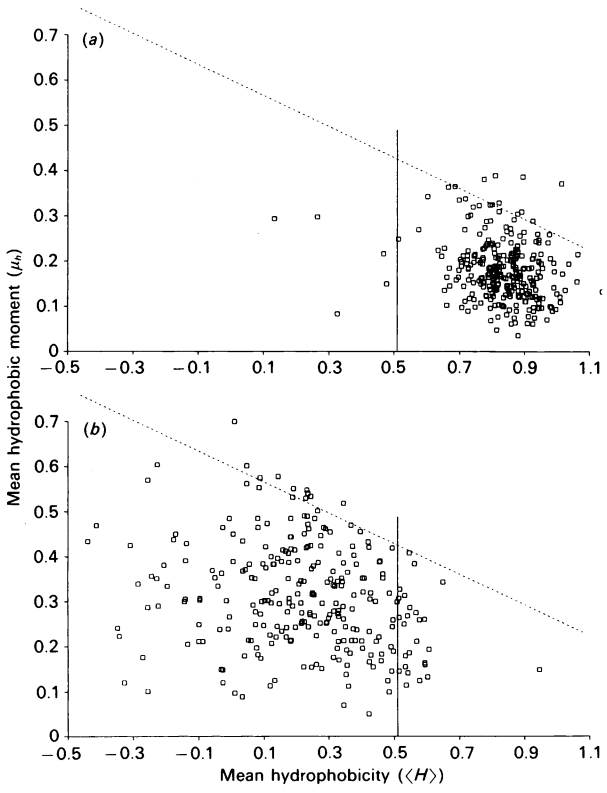

Fig. 10. Scatter plot of mean hydrophobic moment (μ_H) and mean hydrophobicity per residue (⟨*H*⟩) for signal peptides (*a*) and nascent peptides (*b*)

The vertical line divides the peptides as globular type to its right and membrane type to its left. The peptides above the upper line (------) are classified as surface-seeking peptides. The slanted line on the right (·····) further demarcates the membrane-type peptides as transmembrane type (to the right of the line) and multimeric membrane type (to the left of the line).

(1984*a*,*b*) is highly useful for inference of the membrane association of peptides. The mean hydrophobic moment and mean hydrophobicity of the signal sequences (Fig. 10*a*) and nascent peptides (Fig. 10*b*) of equal length to signal sequence are calculated. The average hydrophobic moment $\langle \mu_H \rangle$ was plotted on the ordinate and the mean hydrophobicity per residue $\langle H \rangle$ was plotted on the abscissa. The vertical line in the Figures divides the peptides as globular type to its right and membrane type to its left. The peptides above the upper line (------) are classified as surface-seeking peptides, in accordance with Eisenberg *et al.* (1984*b*). The slanted line on the right (·····) further categorizes the peptides as transmembrane types (to the right of the line) and multimeric membrane-bound types (to the left of the line). Most of the signal sequences are found on the border between monomeric and multimeric membrane-bound types. This suggests that signal peptides cannot be strictly classified in either of these categories. However, their hydrophobic stability within the membrane is substantiated from this plot. Fig. 10(*b*) clearly indicates the preference of nascent peptides to be of globular type. It should be pointed out that very few nascent peptides are classified as surface-seeking. In contrast, signal peptides show their specific membrane-type characteristics. A monomeric peptide has enough free energy to be membrane-bound, whereas a multimeric peptide, without enough free energy of its own to be membrane-bound, derives additional free energy by association with other peptides to be membrane-bound. From Fig. 10(*a*) most of the signal peptides lie in the multimeric regions, indicating that signal peptides may not have sufficient free energy for penetration through membrane, in comparison with known monomeric peptides. However, these signal peptides could associate with the adjacent peptides to form stable multimeric membrane-bound-type peptides. Nevertheless nascent peptides, which prefer to be populated in the globular region (Fig. 10*b*), may not prefer to interact and increase the stability of signal peptides in membranes by their association. This analysis shows that signal peptides could be classified as membrane-preferring peptides but not membrane-stable peptides.

CONCLUSIONS

These results show that the general sequence arrangement of signal peptides has a significant correlation with the properties of the amino acid residues. The presence of the hydrophobic core in the signal peptides is revealed from this study. This study also shows that signal peptides contain the following potential features for recognition: (i) a preference for a basic amino acid at a particular position; (ii) equal conformational preference for α-helix and β-sheet structures in the hydrophobic regions; (iii) specific recognition features at the cleavage site in terms of bulkiness, M_r and acidic charge preference. The preferential occupation of the amino acid residues is shown in conformity with their properties. The hydrophobic moment calculations reveal the possibility of different stable structures. This study further shows that signal peptides prefer to be stabilized within the membrane by association with other peptides. As the subsequent mature chains prefer to be globular, the possibility of their association within the membrane is less probable. Two statistical parameters are derived in these calculations. The standard deviations in each plot show the deviation from mean, and the coefficient of variation is used to compare the deviation in the distribution of different properties.

In designing a signal peptide, the following features should be given adequate consideration: (1) central hydrophobic core; (2) equal preference for α-helix and β-sheet secondary-structure arrangements; (3) bulkiness and M_r of residues at the cleavage site and their preference for turn formation; (4) basic residue at the *N*-terminus; (5) extending the length of the signal peptide with the inclusion of an acidic residue after the cleavage site.

In future, the consensus signal peptides can be designed by using the above rules and Table 3 of preferred residues. These peptides can be modelled from computer graphics and their conformational preferences can be studied by molecular mechanics calculations in both aqueous and lipid environments.

I am grateful to Professor Jere P. Segrest for guidance and support, to Christie G. Brouillette for helpful suggestions and to Professor C. P. Woodbury, Jr., for an in-depth review.

REFERENCES

Blobel, G. & Dobberstein, B. (1975*a*) J. Cell Biol. **67**, 835–851
Blobel, G. & Dobberstein, B. (1975*b*) J. Cell Biol. **67**, 852–862
Briggs, M. S. & Gierasch, L. M. (1986) Adv. Protein Chem. **38**, 119–181
Creighton, T. E. (1983) Protein Structure and Molecular Properties, p. 235, W. H. Freeman, San Francisco
Eisenberg, D., Weiss, R. M. & Terwilliger, T. C. (1984*a*) Proc. Natl. Acad. Sci. U.S.A. **81**, 140–144
Eisenberg, D., Schwary, E., Komaromy, M. X. & Wall, R. (1984*b*) J. Mol. Biol. **179**, 125–142
Engelman, D. M., Steitz, T. A. & Coldman, A. (1986) Annu. Rev. Biophys. Chem. **15**, 321–353
Inouye, M. & Halegoua, S. (1980) CRC Crit. Rev. Biochem. **7**, 339–371
Nesmeyanova, M. A. (1982) FEBS Lett **142**, 189–193
Pearlman, D. & Halvorson, H. O. (1983) J. Mol. Biol. **167**, 391–409
Pincus, M. R. & Klausner, R. D. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 3413–3417
Randall, L. L. & Hardy, S. J. S. (1984) Mod. Cell Biol. **3**, 1–21
Randall, L. L. & Hardy, S. J. S. (1989) Science **243**, 1156–1159
Sokal, R. R. & Rohlf, J. (1969) Biometry, W. H. Freeman, New York
von Heijne G. (1981) Eur. J. Biochem. **121**, 275–278
von Heijne, G. (1984) J. Mol. Biol. **173**, 243–251
von Heijne, G. (1985) J. Mol. Biol. **184**, 99–115
von Heijne, G. & Blomberg, C. (1979) Eur. J. Biochem. **97**, 175–181
Walter, P. & Lingappa, V. R. (1986) Annu. Rev. Cell Biol. **2**, 499–516
Wickner, W. (1981) Science **22**, 861–868
Zimmerman, J. M., Eliezer, N. E. & Simha, R. (1968) J. Theor. Biol. **21**, 170–201

Received 27 July 1989/12 February 1990; accepted 14 February 1990