Eur. J. Biochem. *116*, 419–422 (1981)

On the Hydrophobic Nature of Signal Sequences

Gunnar VON HEIJNE

Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor

(Received October 8, 1980/January 23, 1981)

A number of signal sequences, prokaryotic as well as eukaryotic, have been analyzed in terms of gross amino acid composition and hydrophobicity. It is shown that the amino acid composition of the hydrophic core can be well reproduced in a computer simulation of signal sequence 'evolution' with selection operating on the mean hydrophobicity of the sequence and the non-occurrence of charged residues. The calculated hydrophobicities are interpreted in terms of a model in which the hydrophobic part of the signal sequence partitions directly into the membrane interior, thereby making further translocation of the growing nascent chain possible.

Secreted and trans-membrane proteins are generally translocated across the membrane barrier co-translationally, in a process somehow initiated by a transient N-terminal signal sequence some 15–20-residues long [1]. The physical basis for the signal function is still not clear, but in view of the facts that (a) all signal sequences have a highly hydrophobic central part, and (b) no obvious sequence homologies exist among them, a reasonable working hypothesis is that the most important restriction on these sequences is that they must retain a certain minimum hydrophobicity. Indeed, incorporation of a relatively polar leucine analog *in vitro* into the leucine-rich signal sequences of rat preprolactin and human placental prelactogen has been shown to prevent translocation of these proteins, whereas the leucine-poor signal sequence of the pre-α subunit of human chorionic gonadotropin still initiates translocation under these conditions [2].

In this paper it is shown that the overall amino acid composition of the hydrophobic cores of two samples of signal sequences can be fairly well reproduced in computer-generated samples of 'signal-like' peptides, selected according to two criteria: (a) no charged residues (Asp, Glu, Lys, Arg) are allowed, and (b) each sequence has to have a certain minimum hydrophobicity (see below). Estimates of the free energy gained upon inserting a signal sequence into the non-polar membrane interior further suggests a possible scheme for the translocation-initiating step.

RESULTS

Hydrophobic Free Energy of the Interaction between the Signal Sequence and the Membrane

In an earlier paper [3] we presented estimates of the standard free energy of transfer of a given type of residue from a random coil conformation in water to a helical conformation in a non-polar environment. This allowed a calculation of the differences in peptide-membrane interaction free energy between different parts of a nascent polypeptide chain during the course of translocation.

When studying the (presumed) interaction between signal sequence and membrane, however, we need an absolute value for the binding free energy, and thus a knowledge of the actual secondary structure adopted by the signal sequence in the two environments. Unfortunately, data relevant to this problem is scarce and open to different interpretations. Circular dichroic analysis of a synthetic peptide representing the precursor-specific N-terminal extension of preproparathyroid hormone (including the signal sequence) has indicated that this 30-residue fragment is highly structured in aqueous as well as non-polar media, with 27% α-helix and 43% β-structure in the aqueous buffer, and 46% α-helix (no β) in the non-polar solvent [4]. It was not stated, however, whether the β-structure in water was intermolecular or intramolecular. Furthermore, the signal-like membrane-binding C-terminal peptide in cytochrome b_5 has been shown to be predominantly helical in both media [5]. Melittin, another membrane-binding peptide with hydrophobic characteristics similar to the signal sequences, appears to be in an extended conformation in water (where it forms a tetramer) and essentially helical under 'membrane-like' conditions [6].

Applications of secondary structure prediction schemes to known signal sequences have yielded somewhat conflicting results [7,8], but they at least suggest that these sequences have high potential for forming secondary structure (α or β).

In what follows, we tentatively assume that signal sequences are helical in both aqueous and non-polar environments. Noting that a residue gains on the average 8 kJ/mol of hydrophobic free energy upon forming a helix in water [9], and subtracting this value from our earlier estimates [3], we arrive at an estimate for the standard free energy of transfer of a given type of residue in a helix in water to a helix in a non-polar phase (Table 1). Two additional changes have been made in Table 1, as compared with its earlier version. First, we do not include any 'H-bond' or 'Charge' contributions for Cys and Met, since sulphur-containing groups generally form only weak hydrogen bonds of low energy [9a]. Second, Tyr was earlier erroneously assigned a 'Charge' contribution; however, hydrogen bonds involving the Tyr-OH group are unusually strong [9a], and the 'H-bond' contribution has tentatively been put equal to 21.0 kJ/mol.

The total hydrophobic contribution to the standard free energy of the interaction between signal sequence and membrane can now be calculated for any given sequence simply as the sum of the individual residue contributions (Table 2). As has already been pointed out [10,11], charged and polar side-chain groups are expected to remain outside the membrane as long as the α-carbon of the residue is no more than

Table 1. *Estimated free energies of transfer of residues originally in a helix in water to a helix in a non-polar phase lacking hydrogen-bonding capacity* The H-bond and charge contributions are those from broken hydrogen bonds, and the free energy needed to neutralize a charged residue, respectively

Residue	Hydrophobic contribution	H-bond contribution	Charge contribution	Total
	kJ/mol			
Leu	− 10.1	−	−	− 10.1
Ala	− 4.2	−	−	− 4.2
Val	− 8.4	−	−	− 8.4
Phe	− 14.2	−	−	− 14.2
Ser	− 4.2	10.5	−	6.3
Ile	− 10.5	−	−	− 10.5
Pro	− 7.1	21.0	−	13.9
Gly	0.0	−	−	0.0
Cys	− 6.3	−	−	− 6.3
Met	− 11.3	−	−	− 11.3
Thr	− 6.7	10.5	−	3.8
Tyr	− 16.3	21.0	−	4.7
Gln	− 10.9	21.0	−	10.1
Trp	− 18.9	10.5	−	− 8.4
Asn	− 8.8	21.0	−	12.2
His	− 12.6	21.0	5.9	14.3
Lys	− 13.0	10.5	20.1	17.6
Arg	− 15.5	31.4	31.4	47.3
Glu	− 12.2	21.0	15.9	24.7
Asp	− 8.0	21.0	18.0	31.0

k positions into the membrane-bound helix, with $k = 4$ for Lys, $k = 3$ for Arg and Tyr, $k = 2$ for Glu, Gln, Met and Trp, and $k = 1$ for Asp, Asn, His, Cys, Ser and Thr. In such cases, only the hydrophobic term (first column in Table 1) is taken into account.

Simulation of the Amino Acid Composition of the Signal Sequence

As a simple test of the idea that the only major restrictions placed upon the hydrophobic core of the signal sequence are that it must be uncharged and that its over-all hydrophobicity must be greater than a certain minimum value in order not to impair its function in the initial steps of translocation, we have compared the overall amino acid composition of two samples, one prokaryotic and one eukaryotic, of signal sequence cores (defined as those residues encompassed between the last charged residue of the N terminus and either the first charged residue toward the C terminus or the last residue of the signal, since it is known that this last residue in the signal sequence is not chosen at random [12]), with that of computer-generated counterparts obtained as follows.

Initially, the computer memory contains 50 copies of a poly(U) 'mRNA', coding for a poly(Phe) peptide, 20 codons long. Then, single base substitutions are made at random, and the hydrophobic free energies of the resulting 'peptides' are evaluated using the values given in Table 1. If either of the two criteria listed in the introduction, i.e. (a) no charged residues are allowed, and (b) the hydrophobicity must be greater than a certain threshold value, is violated for a given 'peptide', a new mutation is made in the original 'mRNA' and so on, until an 'mRNA' coding for an allowed 'peptide' is obtained. This new 'mRNA' is then substituted for the original one in the computer memory.

Every 200 rounds of this mutation-selection procedure the overall amino acid composition of the 50 sequences in the memory is calculated and stored, and after a sufficient number of rounds (about 1200) the mean amino acid frequencies of the whole sample are calculated. The mean amino acid frequencies, then are based on 3000 sequences generated as described.

Simulated and observed amino acid distributions are presented in Table 3. The threshold value (criterion b) was chosen so that the mean hydrophobicity per residue was the same in the simulated and observed distributions.

Table 2. *Total hydrophobic contribution to the free energy of the interaction between signal sequence and membrane for some signal sequences*

Protein	Membrane binding segment	Hydrophobic free energy	Reference
		kJ/mol	
Eukaryotic proteins			
Human placental lactogen	3 – 22	−127	[20]
Rat growth hormone	9 – 20	−110	[21]
Bovine parathyroid hormone	6 – 23	−170	[22]
Hen egg white ovomucoid	2 – 18	−132	[23]
Rat insulin	3 – 17	−151	[24]
Mouse immunoglobulin $\varkappa$ chain	2 – 22	−117	[25]
Mouse immunoglobulin λ_1 chain	1 – 19	− 92	[25]
Rat liver albumin	1 – 21	−140	[26]
Dog trypsinogen	1 – 17	−161	[26]
Chick lysozyme	1 – 22	−160	[26]
Human chorionic gonadotropin, α subunit	3 – 20	−171	[27]
Rat lactin	6 – 26	−152	[28]
Rabbit uteroglobin	1 – 17	−135	[29]
Hen egg white conalbumin	1 – 20	−137	[30]
Vs virus G protein	1 – 13	−121	[31]
Mouse immunoglobulin H chain	1 – 13	− 81	[32]
Influenza A hemagglutinin (FPV)	3 – 14	−105	[33]
Influenza A hemagglutinin (Victoria)	1 – 18	−133	[33]
Influenza A hemagglutinin (Jap)	1 – 16	−124	[33]
Avian apoVLDL-II	2 – 15	−133	[34]
Human leukocyte interferon	5 – 18	−112	[35]
Human fibroblast interferon	3 – 15	−122	[35]
Bovine cytochrome b_5	107 – 127	−122	[36]
Prokaryotic proteins			
E. coli lipoprotein	2 – 22	−113	[12]
E. coli ompA protein	1 – 20	−134	[37]
E. coli maltose-binding protein	6 – 30	−151	[18]
E. coli λ receptor	4 – 23	−118	[19]
β-Lactamase pBR 322	5 – 24	−156	[38]
E. coli leucine-binding protein	4 – 19	−102	[39]
M13 phage coat protein (major)	5 – 26	−115	[40]
M13 phage coat protein (minor)	1 – 17	−124	[12]
E. coli chromosomal β-lactamase	1 – 20	− 95	– [a]
E. coli Leu-Ile-Val binding protein	3 – 25	−146	[41]

[a] B. Jaurin, T. Grundström, T. Edlund, and S. Normark, personal communication (1980).

DISCUSSION

As is clear from Table 3, the simulated and observed distributions generally agree quite well (* indicates a difference

Table 3. *Observed and simulated amino acid distributions*
In the signal sequences analyzed (the same as in Table 2), only those residues lying between the last charged N-terminal residue and either the first charged C-terminal residue or the last residue before the peptidase cleavage site (whichever comes first) were included. In each case, the minimum hydrophobicity threshold used in the simulation was chosen so that the mean hydrophobicity per residue was the same in the observed and simulated distributions. * Indicates a difference significant on the 10% level, as judged by comparison with a binominal distribution. For eukaryotic signals, the mean hydrophobicity per residue = −5.2 kJ/mol, for prokaryotic signals the mean hydrophobicity per residue = −3.9 kJ/mol

Residue	Eukaryotic signals		Prokaryotic signals	
	observed	simulated	observed	simulated
	residues/molecule			
Leu	100*	67.7	29	32.4
Ala	40	31.8	44*	16.7
Val	29*	41.4	20	20.8
Phe	30	28.7	12	12.9
Ser	22	22.9	14	15.1
Ile	26	34.2	12	16.4
Pro	13	8.9	7	6.8
Gly	8*	23.9	9	14.4
Cys	14	17.8	6	9.4
Met	12	12.7	8	5.5
Thr	19	18.1	13	11.6
Tyr	6	8.6	1*	5.9
Gln	8	5.8	3	4.0
Trp	6	10.6	0*	5.2
Asn	5	5.1	4	3.7
His	4	4.1	2	3.3
Total	342	342.3	184	184.1

significant on a conservative 10% level). In the prokaryotic sample alanine is the only major exception; possible explanations could be that alanine might be preferentially utilized because it is a very 'cheap' amino acid to synthesize [13], or that it is a strong helix-former [14]. Alanine is also the most abundant amino acid in total *Escherichia coli* protein [15]. In the eukaryotic sample, leucine is particularly abundant (strong helix-former), whereas glycine (strong helix-breaker) is rare.

Minor differences could also be due to, for example, the small sample sizes, to inaccuracies in the estimate in Table 1, or to particular codon usage patterns (in the simulation, all codons corresponding to the same amino acid are given equal weights). To the extent that the precise conformation of the signal sequence is important, additional constraints not accounted for in the simulation would be expected to influence the results somewhat.

The good correspondence thus obtained lends support to the idea that the signal sequence at some point during the initiation of translocation partitions directly into the membrane interior, rather than binds to a 'signal receptor' protein' with a specific recognition site. We [3], as well as others [12], have argued that the signal sequence should bind to the membrane in a U-shaped fashion, with the charged N terminus remaining on the cytoplasmic side. Indeed, cytochrome b_5 seems to bind in this manner to 'non-leaky' lipid vesicles [16], the binding being essentially diffusion-controlled [17].

The minimum number of hydrogen bonds that must be broken when a signal sequence thus binds to a membrane has been estimated to be 8–10 [11], corresponding to an unfavorable free energy contribution of some 85–105 kJ/mol. This value compares well with the total hydrophobic free energies in Table 2, indicating that signal sequences may have been selected on a criterion of net free energy gain when inserted into a membrane. The importance of an uninterrupted and sufficiently long stretch of non-polar residues has recently been demonstrated in signal sequence mutants [18,19].

Lastly, our results suggest a 'passive' role for the signal sequence: rather than actively guiding the ribosome to a translocation site on the membrane, it may be designed simply to keep itself 'out of the way' by partitioning into the bilayer during those instances when the ribosome happens to be close to the membrane, so as not to obstruct the putative binding site on the ribosome and interfere with its subsequent binding to a translocation site.

REFERENCES

1. Blobel, G. & Dobberstein, B. (1975) *J. Cell Biol. 67*, 835–851.
2. Hortin, G. & Boime, I. (1980) *Proc. Natl Acad. Sci. USA, 77*, 1356–1360.
3. von Heijne, G. & Blomberg, C. (1979) *Eur. J. Biochem. 97*, 175–181.
4. Rosenblatt, M., Beaudette, N. V. & Fasman, G. D. (1980) *Proc. Natl Acad. Sci. USA, 77*, 3983–3987.
5. Dailey, H. A. & Strittmatter, P. (1978) *J. Biol. Chem. 253*, 8203–8209.
6. Dawson, C. R., Drake, A. F., Helliwell, J. & Hider, R. C. (1978) *Biochim. Biophys. Acta, 510*, 75–86.
7. Austen, B. M. (1979) *FEBS Lett. 103*, 308–313.
8. Garnier, J., Gaye, P., Mercier, J.-C. & Robson, B. (1980) *Biochimie (Paris) 62*, 231–239.
9. Chothia, C. (1976) *J. Mol. Biol. 105*, 1–14.
9a. Schultz, G. E. & Schirmer, R. H. (1979) *Principles of Protein Structure*, Springer Verlag, New York.
10. Tanford, C. (1978) *Science (Wash. DC) 200*, 1012–1018.
11. von Heijne, G. (1980) *Eur. J. Biochem. 103*, 431–438.
12. Inouye, M. & Halegoua, S. (1979) *Crit. Rev. Biochem. 7*, 339–371.
13. Atkinson, D. E. (1977) in *Cellular Energy Metabolism and Its Regulation*, p. 74, Academic Press, New York.
14. Chou, P. & Fasman, G. D. (1978) *Annu. Rev. Biochem. 47*, 251–276.
15. Lehniger, A. L. (1975) *Biochemistry*, 2nd edn, Worth, New York.
16. Enoch, H. G., Fleming, P. J. & Strittmatter, P. (1979) *J. Biol. Chem. 254*, 6483–6488.
17. Leto, T. L. & Holloway, P. W. (1979) *J. Biol. Chem. 254*, 5015–5019.
18. Bedoulle, H., Bassford, P. J., Fowler, A. V., Zabin, I., Beckwith, J. & Hofnung, M. (1980) *Nature (Lond.) 285*, 78–81.
19. Emr, S., Hedgpeth, J., Clément, J.-M., Silhavy, T. J. & Hofnung, M. (1980) *Nature (Lond.) 285*, 82–85.
20. Sherwood, L. M., Burstein, Y. & Schechter, I. (1979) *Proc. Natl Acad. Sci. USA, 76*, 3819–3823.
21. Seeburg, P. H., Shine, J., Martial, J. A., Baxter, J. D. & Goodman, H. M. (1977) *Nature (Lond.) 270*, 486–494.
22. Rosenblatt, M., Habener, J. F., Tyler, G. A., Sehpard, G. L. & Potts, J. T. (1979) *J. Biol. Chem. 254*, 1414–1421.
23. Thibodeau, S. N., Palmiter, R. D. & Walsh, K. A. (1978) *J. Biol. Chem. 253*, 9018–9023.
24. Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naber, S. P., Chick, W. L. & Gilbert, W. (1978) *Proc. Natl Acad. Sci. USA, 75*, 3727–3731.
25. Burstein, Y. & Schechter, I. (1977) *Proc. Natl Acad. Sci. USA, 74*, 716–720.
26. Strauss, A. W., Bennett, C. D., Donohue, M., Rodkey, J. A. & Alberts, A. W. (1977) *J. Biol. Chem. 252*, 6846–6855.
27. Fiddes, J. C. & Goodman, H. M. (1979) *Nature (Lond.) 281*, 351–356.

28. McKean, D. J. & Maurer, R. A. (1978) *Biochemistry*, *17*, 5215–5219.
29. Malsky, M. L., Bullock, D. W., Willard, J. J. & Ward, D. N. (1979) *J. Biol. Chem. 254*, 1580–1585.
30. Thibodeau, S. N., Lee, D. C. & Palmiter, R. D. (1978) *J. Biol. Chem. 253*, 3771–3774.
31. Lingappa, V. R., Katz, F. N., Lodish, H. F. & Blobel, G. (1978) *J. Biol. Chem. 253*, 8667–8670.
32. Jilka, R. L. & Pestka, S. (1979) *J. Biol. Chem. 254*, 9270–9276.
33. Min Jou, W., Verhoeyen, M., Devos, R., Saman, E., Fang, R., Huylebroeck, D., Fiers, W., Threlfall, G., Barber, C., Carey, N. & Emtage, S. (1980) *Cell*, *19*, 683–696.
34. Chan, L., Bradley, W. A. & Means, A. R. (1980) *J. Biol. Chem. 255*, 10060–10063.
35. Taniguchi, T., Mantei, N., Schwarzstein, M., Nagata, S., Muramatsu, M. & Weissmann, C. (1980) *Nature (Lond.) 285*, 547–549.
36. Fleming, P. J., Dailey, H. A., Corcoran, D. & Strittmatter, P. (1978) *J. Biol. Chem. 253*, 5369–5372.
37. Movva, N. R., Nakamura, K. & Inouye, M. (1980) *J. Biol. Chem. 255*, 27–29.
38. Sutcliffe, J. G. (1978) *Proc. Natl Acad. Sci. USA*, *75*, 3737–3741.
39. Oxender, D. L., Anderson, J. J., Daniels, C. J., Landick, R., Gunsalus, R. P., Zurawski, G. & Yanofsky, C. (1980) *Proc. Natl Acad. Sci. USA*, *77*, 2005–2009.
40. Sugimoto, K., Sugisaki, H., Okamoto, T. & Takanami, M. (1977) *J. Mol. Biol. 110*, 487–507.
41. Oxender, D. & Landick, B. (1981) in *Membranes and Transport: a Critical Review* (Martonosi, A., ed) Plenum Press, New York, in the press.

G. von Heijne, Forskingsgruppen för teoretisk biofysik, Institutionen för teoretisk fysik, Kungliga Tekniska Högskola, S-100 44 Stockholm, Sweden