Pharmaceutical Research, Vol. 6, No. 9, 1989

Research Article

Amino Acids Bracketing the Predicted Transmembrane Domains of Membrane Proteins

Charles Pidgeon,[1,2] Renee L. Williard,[1] and Susan C. Schroeder[1]

Received December 21, 1988; accepted April 3, 1989

The cell membrane is a complex mixture of several classes of biomolecules but amino acids and lipids are the main constituents. For this reason we are establishing a data base of transmembrane proteins with the intent of using the data base to identify interfacial peptide sequences useful for studying protein–lipid interactions at membrane interfaces. Our present intention is to characterize transmembrane peptides and amino acids found near the membrane interface. A data base containing only signal peptides is available (G. von Heijne. *Prot. Seq. Data Anal.* 1:41–42, 1987).

KEY WORDS: interfacial peptides; membrane peptides: protein–lipid interactions.

INTRODUCTION

The diversity of biological functions associated with different membrane proteins justifies the study of protein–lipid interactions near the interfacial region of cell membranes. Membrane proteins have several functions including ion transport (1–4), anchoring proteins to the cell surface, translocating information across the cell membrane (5,6), and drug resistance (7–9). However, functional membrane proteins are not always synthesized as mature proteins; some membrane proteins require posttranslational processing for normal biological activity. For example, HTLV viruses (10), paramyxoviruses (11), and orthomyxoviruses (12) express fusogenic membrane proteins that require posttranslational proteolytic cleavage for normal biological activity. In addition to processing membrane proteins by posttranslational proteolytic cleavage, some membrane proteins are processed by the attachment of fatty acids co- or posttranslationally. Fatty acid acylation of proteins participates in controlling the intracellular distribution of the molecule. The majority of palmitoylated proteins is anchored to the cyto-

[1] Department of Industrial and Physical Pharmacy, School of Pharmacy and Pharmacal Sciences, Purdue University, West Lafayette, Indiana 47907.

[2] To whom correspondence should be addressed at Department of Medicinal Chemistry and Pharmacognosy, Purdue University, West Lafayette, Indiana 47907.

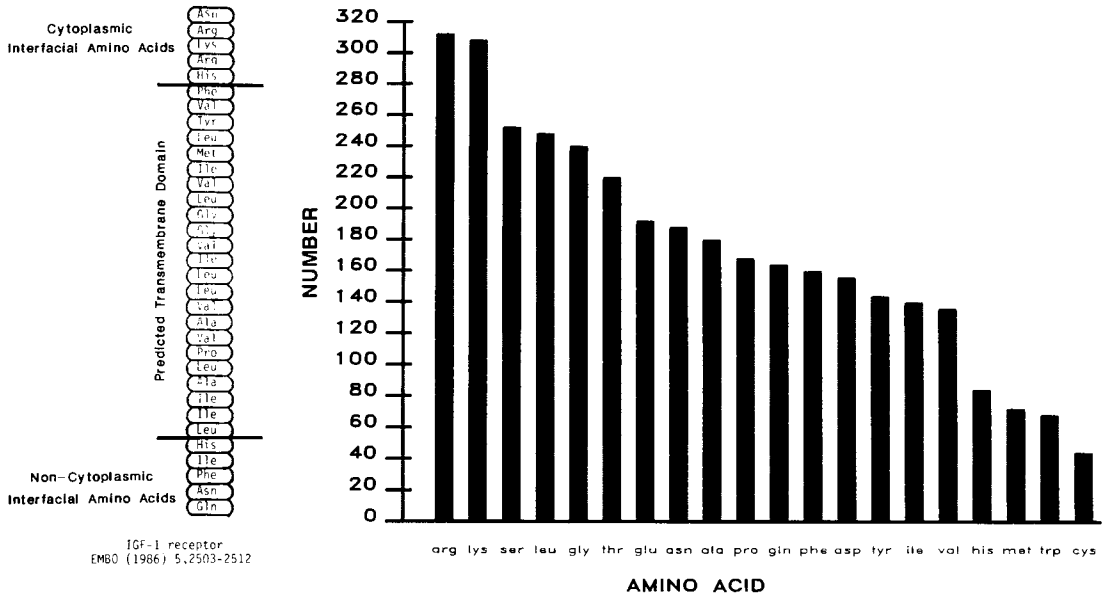

Fig. 1. Distribution of amino acids at the membrane interface. Both cytoplasmic and noncytoplasmic interfacial amino acids are included in the frequency distribution. The method of data tabulation using the IGF-1 receptor is shown.

0724-8741/89/0900-0779$06.00/0

Table I. Proteins Containing Transmembrane Sequences

	Ref No.
Acetylcholine receptor	
T. californica	108
T. marmorata	55
Muscarinic, porcine, cerebrum	87
Muscarinic, porcine, atrial	110
Nicotinic, *Drosophila*	78
Adrenergic receptor	
B_2, hamster	56
B, avian	148
B_2, human	57
ADP/ATP translocator protein	
N. crassa	34
S. cereviseae	34
Bovine	34
Asialoglycoprotein receptor	
Human, liver	129
Rat, liver	82
Aspartate receptor	
S. typhimurium	119
ATPase	
H^+ (*N. crassa*)	74
H^+ (*E. coli*)	79, 140
K^+ (*E. coli*)	80
Ca^{2+} (sarcoplasic reticulum)	80
Ca^{2+} + Mg^{2+} (rabbit, muscle)	95
Yeast plasma membrane	149
H^+ + K^+ (rat stomach)	126
Na^+ + K^+ beta subunit (sheep/kidney)	127
Na^+ + K^+ alpha subunit (sheep/kidney)	129
Bacteriorhodopsin (*H. halobium*)	60
Beta protein of amyloid fiber	113
CAM (cell adhesion molecules)	
L-CAM (liver cell adhesion, chick)	68
N-CAM (neuronal cell adhesion, chick)	77
Can 1 (amino acid permease, *S. cerevisiae*)	81
Cytochrome p-450	
p450 isozyme 2 rabbit liver	133
p450 porcine, hepatic	76
Cytochromes (general)	
Cytochrome b of complex III yeast	142
Cytochrome b_2 (yeast)	73
Cytochrome b5 (bovine, liver)	64, 53
Cytochrome *c* peroxidase	86
Enkakephalinase	54
Epidermal growth factor receptor (human)	5
Fc receptors	
IgG murine (Fc-gamma, R-alpha)	112
IgG murine (Fc-gamma, R-beta-1)	112
IgG murine (Fc-gamma, R-beta-2)	112
IgE human (Fc-epsilon)	84
Fibronectin receptor	131
Glucose transporter protein (human, hepatic)	102
Glutamyl transpeptidase (rat, kidney)	89
Glycophorin, erythrocyte (human)	134
Hemolysin B (*E. coli*)	61
Histocompatibility antigens	
PH-2^d-1 (H-2 mouse)	44
PH-2^d-3 (H-2 mouse)	44
H-$2K^b$ (mouse, class I)	48
I-AK (class II, mouse)	41
I-AD (class II, mouse)	41
I-EK (class II, mouse)	98
MHC II alpha-chain (human)	93

Table I. Continued

	Ref No.
MHC II HLA S alpha	38
MHC II beta-chain (human)	150
HLA-DR (human) invariant chain (alpha)	130
HLA-DZ (alpha, human)	135
HLA-DP (alpha, human)	38
HLA-DQW-1 (alpha, human)	38
HLA-DQW-2 (alpha, human)	46
HLA-DX (alpha, human)	38
HLA-DR (alpha, human)	93
HLA-DC (human)	
Alpha (invariant chain, 2 cell lines)	
PIIa5	122
PDCHl	122
Beta (variant chain, 3 cell lines)	
PII-B-1	122
PII-B-2	122
COSII-102	122
Immunoglobulins	
IgG1, IgG2a, IgG2b mouse, gamma chain	144
IgM (mu chain nonsecreted form)	115
Immunoglobulin receptors	
IgA and IgM (rabbit)	101
Insulin receptor (human)	22
Insulin-like growth factor receptor (IGF-1)	6
Integrin (chick embryo fibroblasts)	132
Interleukin-2 receptor (human)	49
Lactose permease	139
Leader peptidase (*E. coli*)	151
Lectins	
Hepatic, chicken (clears blood proteins)	58
Lens fiber (bovine)	71
Leukocyte common antigen	
Human	111
Rat	39
Mouse	120
LDL receptor (bovine, adrenal)	118
LDL receptor (human)	146
Maltose transport proteins	
malF (*E. coli*)	66
Melibiose carrier (*E. coli*)	147
MDR cell surface antigen	
Mouse	72
Human MDR 1	8
Hamster	9
MRCOX-2 antigen (mouse)	47
Myelin-associated proteins	
Myelin proteolipid, bovine, brain	90
Myelin-associated glycoprotein	37
Nerve growth factor receptor	
Human	85
Oncogenes/protooncogenes	
Neu (human)	40
Neu (rat)	26
HER 2	27
c-erb-B2 (human)	145
c-fms (proto-oncogene)	50
c-ros-1	99
c-fes/fps (proto-oncogene)	114
v-fms	75
v-erb B (avian)	27
v-ros (ASV)	107
c-ros (chicken)	106
Phosphotransferase (enzyme II^{mtl})	91

Table I. Continued

	Ref No.
Pro-sucrase-isomaltase	83
Ribophorins	
Ribophorin I (human)	52
Ribophorin II (human)	52
Rhodopsin/opsin	
Bovine	104
Halorhodopsin	42
Human	105
RH2 opsin (*Drosophila*)	51
R7 opsin (*Drosophila*)	67
Serine receptor	
E. coli	43
Spumatovirus	
Human	65
T_3/T-cell receptor complex (human)	
Gamma chain	94
Delta chain	136
Epsilon chain	70
T_3/T-cell receptor complex (murine)	
Alpha chain	121
Beta chain	121
Delta chain	137
T cell	
CD2 (T11) antigen	125
Thy 1 (rat)	124
T-cell growth factor (rat)	92
Transferrin receptor (human)	123
Uncoupling protein, mitochondria	36
UDP-glucuronosyltransferases	
$UDPGT_R$-2F	97
$UDPGT_R$-3	96
Virus membrane proteins (host/mammal)	
Adenovirus AD5 (E3 region)	143
Coronovirus (E1, MHV-A59)	116
EBV1	62
Herpes simplex virus (gD)	141
Influenza virus	
M_2 (human)	88
A/FPV/Rostock/34	100
Hong Kong, A/Aichi/2/68	138
Hong Kong A/Victoria/3/75	138
Neuraminidase	63
Rabies glycoprotein (ERA)	35
Semilike Forest virus membrane glycoprotein	69
Vacinnia virus polypeptide	45
Vesicular stomatitis virus	117
Moloney murine leukemia virus	59
G.A.-feline leukemia virus	59
Akv	152
Influenza neuraminidase	
N_2 A/Victoria/3/75	153
N_1 A/PR8/34	153
N_1 A/WSN/33	153
Adenovirus E-19	109
Virus membrane proteins (host/bacteria)	
If1	103
Ike	103
zj2	103
fd(f1)(m13)	103
Vitronectin receptor	131

plasmic side of the plasma membrane, whereas myristoylated proteins may or may not be membrane associated inside the cell (13,14). With regard to protein processing, it is believed that myristoylation occurs cotranslationally, whereas palmitoylation occurs posttranslationally (13,14).

This *diversity* of membrane protein function and the cellular *processing* (cleavage, myristoylation, palmitoylation, etc.) of membrane proteins encouraged us to study protein–lipid interactions at the membrane interface with the intent of trying to identify molecules that will interact with proteins at this region of the cell membrane. Molecules that specifically bind to an interfacial peptide sequence may perturb the function of the membrane protein and/or inhibit processing of the membrane protein. Molecules that affect protein function by acting near the interfacial region of the cell may have biological activity and thus drug leads may evolve from this work.

Our first goal is to identify model compounds for the study of protein–lipid interactions at the membrane interface. Consequently a data base of transmembrane proteins was established and the frequency of amino acids bracketing the predicted transmembrane domains of approximately 400 transmembrane sequences was calculated. Below is a description of the data base and histograms describing the distribution of amino acids near the interfacial region of cell membranes. Concurrently with establishing the data base we are developing high-performance liquid chromatographic (HPLC) techniques to study protein–lipid interactions. The HPLC technique, denoted immobilized artificial membranes, involves covalently bonding membrane lipids to solid supports to study protein–lipid interactions (15).

METHODS

Identification of Transmembrane Peptide Sequences

All transmembrane peptide sequences were obtained from published sequence data. The transmembrane domain of proteins is virtually always predicted from computer algorithms. Predictions are based on energy calculations (16), hydrophobic moment plots (17,18), or hydropathy plots (19). These algorithms adequately predict regions of proteins that span membranes, but the accuracy of these programs in clearly specifying where the membrane spanning domain begins and ends has never been addressed. Because of this uncertainty, the definition of membrane interface was established based on the physicochemical nature of membrane lipids and proteins comprising the membrane as described below.

Definition of the Membrane Interface

The primary difficulty in defining the cell membrane interface and identifying amino acids residing in this region is that the structure of membrane proteins and the position of membrane proteins in the membrane is unknown. Lack of this knowledge makes identifying interfacial amino acids somewhat speculative. Nevertheless, a functional definition of the membrane interface is necessary for tabulating the expected frequency distribution of amino acids bracketing the predicted transmembrane domain of proteins. A functional definition of the membrane interface can be developed

by comparing the size of lipid headgroups to the size of peptide sequences as shown below.

Lipids with different headgroups may protrude from the membrane's glycerohydrocarbon region by as little as 4 Å for layer-parallel phosphatidylcholine (PC) or more than two times this distance (10 Å) as found for lactocerebrosides (20). Solute binding to lipid headgroups can cause conformational changes resulting in different thicknesses of the headgroup region. This is exemplified by extended PC headgroups being 7 Å when lanthanide ions bind (21). Thus different lipid headgroups interact with transmembrane peptides at different distances from the glycerohydrocarbon region of the membrane, depending on the headgroup conformation and headgroup size.

Peptides in an α-helical conformation require 3.6 amino acids per turn and the distance between helical turns is about 5.4 Å. An α-helical peptide exiting the membrane's glycerohydrocarbon region and penetrating the lipid headgroup region would require approximately one helical turn (or four amino acids) to escape layer-parallel PC headgroups (4 Å). As discussed, headgroup conformations vary, and consequently more than four amino acids (for α-helical proteins) will have direct access to lipid headgroups, particularly PC headgroups. We note that the conformation of peptides at the interfacial region of cell membranes is unknown and all peptides may not assume an α-helical conformation.

From these considerations, the interfacial region of the membrane is variable but at least five amino acids extending form the predicted transmembrane domain interact directly with membrane lipid headgroups. These five interfacial amino acids bracketing the predicted transmembrane domains were used as a first approximation of interfacial amino acids.

Data Tabulation

Data were tabulated as depicted in Fig. 1 for the insulin-like growth factor (IGF-1) receptor. The five interfacial amino acids bracketing the predicted transmembrane domain of the IGF-1 receptor are His-Arg-Lys-Arg-Asn and His-Ile-Phe-Asn-Gln on the cytoplasmic and noncytoplasmic side of the membrane, respectively (Fig. 1). Due to the difficulty in unambiguously specifying the cytoplasmic and noncytoplasmic side of all transmembrane sequences in the data bank, data calculation combined amino acids bracketing both sides of predicted transmembrane sequences.

Proteins listed in Table I were used for data calculation. Approximately 150 proteins were used for the analysis. Histograms were calculated from 773 interfacial peptide sequences; however, 632 interfacial amino acid sequences contained exactly "5 amino acids in length." The remaining interfacial sequences were obtained from short connecting sequences linking two transmembrane domains. Short connecting sequences containing less than 10 amino acids between adjacent transmembrane domains did not have enough residues to assign 5 interfacial amino acids to each transmembrane sequence. Consequently, short connecting sequences linking transmembrane peptides were included in the data calculations but the amino acids were partitioned between two transmembrane domains. Only sequence data references in Table I were used, even though some of these references reported only one domain or part of a protein. These proteins are listed in Table I as the alpha, beta, etc., subunit for the proteins in question.

Some proteins contain identical subunits, and consequently identical transmembrane domains. For instance, the acetylcholine receptor and insulin receptor (22) contain identical subunits. In contrast to identical subunits within the same protein, some different proteins share a common subunit. A group of proteins sharing a common subunit are the lymphocyte function-associated antigen (LFA-1), compliment receptor type 3 (CR3), and the p150, 95 antigen. These three proteins are heterodimeric complexes containing a unique α subunit noncovalently associated with a common β subunit (23). These identical transmembrane domains from identical protein subunits were included only once in the calculated frequency distribution.

Statistical Calculations

For statistical calculations, X_i was the observed frequency of amino acid i bracketing all predicted transmembrane domains; P_i was the probability of finding amino acid i in the average protein (based on the Dayhoff data base); N was 3479, the total amino acids bracketing the predicted transmembrane domains listed in Table I; and NP_i was the number of expected occurrences of amino acid i near the membrane interface. Each amino acid residue was treated as an independent population. For each amino acid i the binomial standard deviation $\sigma_i = [NP_i(1 - P_i)]^{1/2}$ was calculated and σ_i was then transformed into a Z scale by $Z = (X_i) - NP_i)/\sigma_i$. Z is thus a standardized scale. We evaluated the statistical significance of each Z score by reference to a standard normal (Gaussian) distribution. (The normal approximation to the binomial distribution is quite good if $NP \geq 5\sigma$; this was true for each amino acid we studied.) The significance threshold chosen was ±3 on the Z scale. This threshold corresponds to the 5% significance level ($\alpha = 0.05$) but conservatively adjusted by the Bonferroni method (24). The Bonferroni adjustment divides α by 20 because there are 20 amino acids under consideration. From tables of the normal distribution we find that the value of Z corresponding to $\alpha = 0.05/20 = 0.0025$ is $Z = \pm 3$.

RESULTS

The distribution of amino acids bracketing the predicted transmembrane domains of proteins in Table I is shown in Fig. 1.[3] The most frequently occurring amino acids in descending order are Arg ≃ Lys > Ser ≃ Leu ≃ Gly. Arg and Lys both had approximately 310 occurrences. Ser, Leu, and Gly had similar occurrences ≃ 225. The least frequently found amino acids at the membrane interface were Cys < Trp ≃ Met < His. This distribution might have been anticipated in view of the function these particular amino acids play in proteins. Cys stabilizes the tertiary structure of proteins; His and Trp are frequently found in the active site of

[3] We have calculated amino acid frequency distributions for the cytoplasmic and noncytoplasmic sides of the cell membrane. However, unambiguous assignment of transmembrane orientation was not possible. Consequently, these distributions have not been included here but are available from the author.

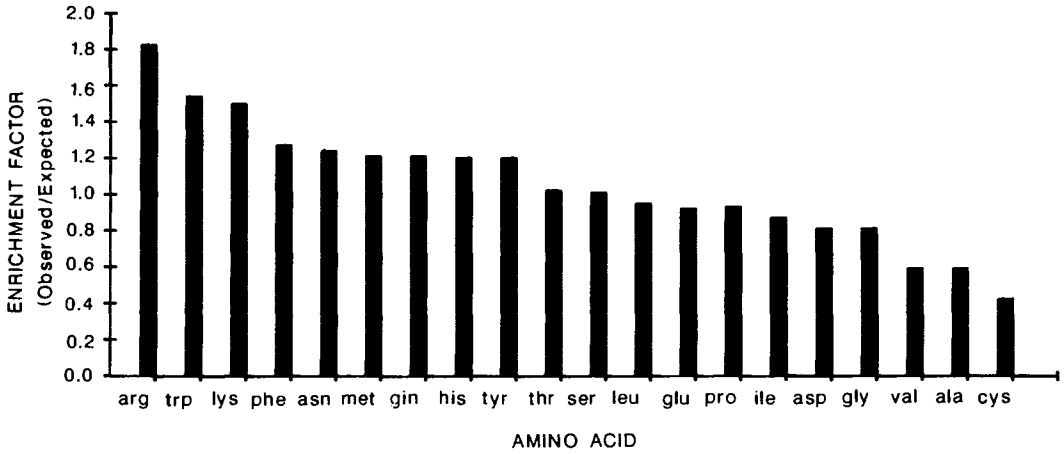

Fig. 2. Enrichment of amino acids at the membrane interface. The distribution was calculated by normalizing the distribution shown in Fig. 1 by the natural abundance of amino acids in proteins.

proteins. Although Cys, Trp, Met, and His are rarely found at the membrane interface, they may in fact play an important role in the function of the protein bearing the group.

Enrichment or depletion of amino acids in the interfacial region can be evaluated by normalizing the distribution by the relative abundance of amino acids found in all proteins. The natural abundance of amino acids in proteins was originally calculated by Dayhoff (25) using 314 proteins; one protein from each class of proteins was used for the distribution. Normalization of the distribution of interfacial amino acids (Fig. 1) for natural abundance using the Dayhoff distribution yields the distribution shown in Fig. 2. Based on Fig. 2, Arg, Trp, and Lys are enriched at the membrane interface, whereas, Val, Ala, and Cys are depleted from this region. The enrichment of Trp at the membrane interface, i.e., the boundary between hydrophobic and hydrophilic regions, is consistent with the hydropathy value of Trp = −0.9 (19). This hydropathy score indicates that Trp is neither hydrophobic nor hydrophilic. Trp thus prefers the interfacial region of protein surfaces and also the interfacial region of cell membranes. It is interesting that Cys is the least likely amino acid at the membrane interface (Fig. 2) and is even depleted relative to the natural abundance of amino acids in proteins (Fig. 2).

The probability of each amino acid bracketing a predicted transmembrane domain was evaluated using the Z statistical scale, as explained above. Figure 3 shows the Z score for each amino acid. Values falling outside the broken parallel lines at $Z = +3$ and $Z = -3$ are unlikely to represent chance variation. Based on Fig. 3, the enrichment of Arg, Lys, Trp, Phe, Asn, and the Depletion of Cys, Ala, Val, Gly, in peptide sequences bracketing predicted transmembrane domains is statistically significant. The depletion of Val and Ala is expected since these nonpolar amino acids are more likely to be found in the hydrophobic domains of proteins and not the interfacial regions of cell membranes.

DISCUSSION

Arg and Lys are frequently found bracketing predicted transmembrane sequences but our histograms did not distinguish amino acids on the cytoplasmic and noncytoplasmic side of the cell membrane. Several studies have noted the occurrence of Arg and Lys on the cytoplasmic side of the membrane (26–29). However, some studies note the occurrence of Arg and Lys *bracketing* transmembrane sequences in a few proteins (26–28), whereas other studies describe the distribution of Arg and Lys in membrane proteins for *entire peptide sequences* found on the cytoplasmic side of the cell membrane (29).

When multiple basic residues are present at the membrane interface of transmembrane peptides, it has been postulated that these charged residues function as a stop-transfer sequence (30). Stop-transfer sequences were postulated to be the recognition sequences other proteins require to abort the secretion of integral membrane proteins. The net result would be to halt protein translocation at the membrane interface leaving the integral membrane protein with the proper transmembrane domain. This hypothesis was recently challenged (31). Using the coliphage f1 gene III protein (pIII), Davis and Model demonstrated that $(Leu\text{-}Ala\text{-}Leu\text{-}Val)_4$ randomly inserted into the pIII gene cause the protein to anchor (i.e. stop transfer) in the membrane wherever this artificial transmembrane peptide was inserted (31).[4] In other words, the membrane topology of the protein was systematically and predictably varied because the artificial membrane spanning domain could anchor the protein in the membrane. From their work, they concluded that the stop-transfer function of peptide sequences reflects hydrophobicity and not some more subtle feature of structure and sequence. These experiments do not support the ideas that multiple basic residues at the membrane interface function only as a stop-transfer sequence for membrane proteins because hydrophobic sequences 16 or more amino acids in length are sufficient to stop transfer. However, it is reasonable to expect basic or charged residues bracketing transmembrane peptide sequences to participate in controlling the exact positioning of transmembrane sequences in cell membranes.

Other experiments by Szczesva-Skorupa *et al.* with sig-

[4] The endogenous transmembrane domain of PIII was deleted for these studied.

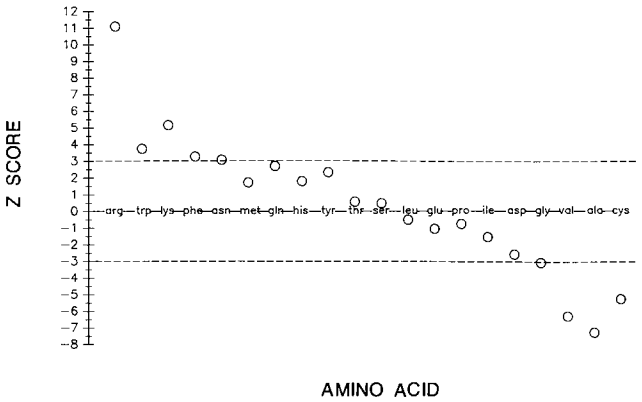

Fig. 3. Calculated Z factor for each amino acid. Amino acids are ordered on the x axis as shown in Fig. 2. Dashed lines denote 3 standard deviations from the mean.

nal peptides demonstrate a functional role for interfacial Arg and Lys (32). Szczesva-Skorupa *et al.* replaced the signal peptide of the secreted hormone preproparathyroid with the hydrophobic N-terminus of cytochrome P-450 and aborted processing and translocation of the preproparathyroid. However, mutagenic replacements of Met-Asp-Leu to Met-Arg-Lys at the N-teminus of the cytochrome P-450 hydrophobic sequence resulted in an artificial signal sequence with normal biological activity. With this artificial signal sequence containing Arg and Lys, preproparathyroid was processed and transported across the endoplasmic reticulum membrane (33). From these experiments it was concluded, "that stop-transfer function is not a simple function of hydrophobic interactions with membrane lipids but may involve interactions with membrane proteins."

In a recent review of protein membrane assembly, Wickner emphasized that signal peptides are vital for membrane assembly but structural features of the mature protein are also important (33). This suggests that the role of basic residues at membrane interfaces is determined by the protein bearing the membrane binding domain. For signal peptides, basic residues permit protein translocation across the membrane but for transmembrane peptides the function of interfacial basic amino acids is uncertain. At the membrane interface, Arg and Lys side chains may bind negatively charged membrane lipids. Interfacial Arg and Lys may thus promote stable association between membrane proteins and the boundary lipids surrounding transmembrane sequences.

In summary we are establishing a data bank to identify biologically relevant sequences to study peptide lipid interactions at the membrane interface. Table I provides sequence data for obtaining these interfacial peptide sequences. The function of amino acids bracketing predicted transmembrane domains is uncertain but most likely depends on the protein bearing the transmembrane sequence.

ACKNOWLEDGMENTS

We thank Sharon Armao for help in preparing the manuscript, Pat White for batch programs to search the MBCRR data base, and Phil Low and Robert Gehland for helpful discussions. We thank Dr. Myra L. Samuels in the statistic department for helpful discussion on the appropriate statistical analysis. This work was supported in part by NIH Grant AI 25712-01.

REFERENCES

1. B. A. Wallace, W. R. Veatch, and E. R. Blout. *Biochemistry* 20:5754–5760 (1981).
2. A. Theretf, J. L. Ranck, and J. F. Tocanne. *BBA* 732:499–508 (1983).
3. N. Roos and D. Broiczka. *BBA* 686:204–214 (1982).
4. M. K. Mathew, R. Nagaraj, and P. Balaram. *J. Biol. Chem.* 257:2170–2176 (1982).
5. A. Ullrich, L. Coussens, T. S. Hayflick, T. J. Dull, A. Gray, A. W. Tam, J. Lee, Y. Yarden, T. A. Libermann, J. Schlessinger, J. Downward, E. L. V. Mayes, N. Whittle, M. D. Waterfield, and P. H. Seeburg. *Nature* 309:418–425 (1984).
6. A. Ullrich, A. Gray, A. W. Tam, T. Yang-Fend, M. Tsubokawa, C. Collins, W. Henzel, T. LeBon, S. Kathuria, E. Chen, S. Jacobs, U. Francke, J. Ramanchandran, and Y. Fujita-Yamaguchi. *EMBO J* 5:2503–2512 (1986).
7. P. Gos, J. Croop, and D. Housman. *Cell* 47:371–380 (1986).
8. C. Chen, J. E. Chin, K. Ueda, D. P. Clark, I. Pastan, M. M. Gottesman, and I. B. Roninson. *Cell* 47:381–389 (1986).
9. J. H. Gerlach, J. A. Endicott, P. F. Juranka, G. Henderson, F. Sarangi, K. L. Deuthars, and V. Ling. *Nature* 324:485–489 (1986).
10. M. Kowalski, J. Potz, L. Basripour, T. Dorfman, W. C. Goh, E. Terwilliger, A. Dayton, G. Rosen, U. Haseltine, and J. Sodroski. *Science* 237:1351–1355 (1987).
11. C. Richardson, D. Hall, P. Greer, K. Hasel, A. Berkovich, G. Englund, W. Bellini, B. Rima, and R. Lazzarini. *Virology* 155:508–523 (1986).
12. M. N. Waxham, A. C. Server, H. M. Goodman, and J. S. Wolinsky. *Virology* 159:381–388 (1987).
13. C. A. Wilcox and E. N. Olson. *Biochemistry* 26:1029–1036 (1987).
14. L. E. Henderson, H. C. Krutzsch, and S. Orostlan. *S. Proc. Natl. Acad. Sci.* 80: 339–343 (1983).
15. C. Pidgeon and U. V. Venkatarum. *Anal. Biochem.* 176:36–47 (1989).
16. R. H. Guy. *Biophys. J.* 45:249–261 (1984).
17. D. Eisenberg. *Annu. Rev. Biochem.* 53:595–623 (1984).
18. D. Eisenberg, R. M. Weiss, and T. C. Terwilliger. *Nature* 299:371–374 (1982).
19. T. Kyte and R. F. Doolittle. *J. Mol. Biol.* 157:105–132 (1982).
20. W. Curatolo, D. M. Small, and G. G. Shipley. *Biochim. Biophys. Acta* 468:11–20 (1987).
21. D. Lichtenberg, S. Amselem, and I. Tamir. *Biochemistry* 18:4169–4172 (1979).
22. A. Ullrich, J. R. Bell, E. Y. Chon, R. Herrera, L. M. Petruzzelli, T. J. Dull, A. Gray, L. Coussens, Y. C. Liao, M. Tsubokawa, A. Mason, P. H. Seeburg, C. Grunfeld, O. M. Rosen, and J. Ramachandran. *Nature* 313:756–761 (1985).
23. S. K. A. Law, J. Gagnon, J. E. K. Hildreth, C. E. Wells, A. G. Willis, and A. J. Wong. *EMBO J* 6:915–919 (1987).
24. M. L. Samuels. *Statistics for the Life Sciences,* Dellen, San Francisco, 1989.
25. M. Dayhoff. *Atlas Protein Seq. Struct.* 5:D-355 (1972).
26. C. I. Bargman, M. Hung, and R. A. Weinberg. *Cell* 45:649–657 (1986).
27. L. Coussens, T. L. Yang-Feng, Y. Liao, E. Chen, A. Gray, J. McGrath, P. H. Seeburg, T. A. Libermann, J. Schlessinger, U. Francke, and A. Louison. *Science* 230:1132–1139 (1985).
28. L. Coussens, C. V. Beueren, D. Smith, E. Chen, R. L. Mitchell, C. M. Isacke, I. M. Verma, and A. Ullrich. *Nature* 320:277–280 (1986).
29. G. von Heijne. *EMBO J.* 5:3021–3027 (1986).
30. G. Blobel. *Proc. Natl. Acad. Sci. USA* 77:1496–1500 (1980).
31. N. G. Davis and P. Model. *Cell* 41:607–614 (1985).
32. E. Szczesna-Skorupa, N. Browne, D. Mead, and B. Kemper. *Proc. Natl. Acad. Sci. USA* 85:738–742 (1988).
33. W. Wickner. *Biochemistry* 27:1081–1086 (1988).
34. G. S. Adrian, M. T. McCammon, D. L. Montgomery, and M. G. Douglas. *Mol. Cell. Biol.* 6:626–634 (1986).

35. A. Anilionis, W. H. Wunner, and P. J. Curtis. *Nature* 294:275–278 (1981).
36. H. Aquila, T. A. Link, and M. Klingenberg. *EMBO J.* 4:2369–2376 (1985).
37. M. Arquint, J. Roder, L. Chia, J. Down, D. Wilkinson, H. Bayley, P. Braun, and R. Dunn. *Proc. Natl. Acad. Sci. USA* 84:600–604 (1986).
38. C. Auffrey, J. W. Lillie, D. Arnot, D. Grossberger, D. Kappes, and J. L. Strominger. *Nature* 308:327–323 (1984).
39. A. N. Barclay, D. I. Jackson, A. C. Willis, and A. F. Williams. *EMBO J.* 6:1259–1264 (1987).
40. C. I. Bargman, M. Hung, and R. A. Weinberg. *Nature* 319:226–230 (1986).
41. C. O. Benoist, D. J. Mathis, M. R. Kanter, V. E. Williams, and H. O. McDevitt. *Cell* 34:169–177 (1983).
42. A. Blanck and D. Osterhelt. *EMBO J.* 6:265–273 (1987).
43. A. Boyd, K. Kendall, and M. I. Simon. *Nature* 301:623–626 (1983).
44. F. Bregegere, J. P. Abastado, S. Kvist, L. Rask, J. L. Lalanne, H. Garoff, B. Cami, K. Wiman, D. Larhammar, P. A. Peterson, G. Gachelin, P. Kourilsky, and B. Dobberstein. *Nature* 292:78–81 (1981).
45. J. P. Brown, D. R. Twardzik, H. Marquardt, and G. J. Todaro. *Nature* 313:491–492 (1985).
46. H. C. Chang, T. Moriuchi, and J. Silver. *Nature* 305:813–815 (1983).
47. M. J. Clark, J. Gegnon, A. F. Williams, and A. N. Barclay. *EMBO J.* 4:113–118 (1985).
48. J. E. Coligan, T. J. Kindt, H. Uehara, J. Martinko, and S. G. Nathenson. *Nature* 291:35–39 (1981).
49. D. Cosman, D. P. Cerretti, A. Larsen, L. Park, C. March, S. Dower, S. Gillis, and D. Urdal. *Nature* 312:768–777 (1984).
50. L. Coussens, C. Van Beveren, D. Smith, E. Chen, R. L. Mitchell, C. M. Isacke, I. M. Verma, and A. Ullrich. *Nature* 320:277–280 (1986).
51. A. F. Cowman, C. S. Zuker, and G. M. Rubin. *Cell* 44:705–710 (1986).
52. C. Crimaudo, M. Hortsch, H. Geusepohl, and D. I. Meyer. *EMBO J.* 6:75–82 (1987).
53. H. A. Dailey and P. Strittmatter. *J. Biol. Chem.* 253:8203–8209 (1978).
54. A. Devault, C. Lazure, C. Nault, H. Le Moual, N. G. Seidah, M. Chretien, P. Kahn, J. Powell, J. Mallet, A. Beaumont, B. P. Roques, P. Crine, and G. Boileau. *EMBO J.* 6:1317–1322 (1987).
55. A. Devillers-Thiery, J. Giraudat, M. Bentaboulet, and J. P. Changeux. *Proc. Natl. Acad. Sci. USA* 80:2067–2071 (1983).
56. R. A. F. Dixon, B. K. Kobilka, D. J. Strader, J. L. Benovic, H. G. Dohlman, T. Frielle, M. A. Bolanowski, C. D. Bennett, E. Rands, R. E. Diehl, R. A. Mumford, E. E. Slater, I. S. Sigal, M. G. Caron, R. J. Lefkowitz, and C. D. Strader. *Nature (London)* 321:75–79 (1986).
57. H. G. Dohlman, M. G. Caron, and R. J. Lefkowitz. *Biochemistry* 26:2657–2664 (1987).
58. K. Drickamer. *J. Biol. Chem.* 256:5827–5839 (1981).
59. J. H. Elder and J. I. Mullins. *J. Virol.* 46:871–880 (1983).
60. D. M. Engelman, R. Henderson, A. D. McLachlan, and B. A. Wallace. *Proc. Natl. Acad. Sci. USA* 77:2023–2027 (1980).
61. T. Felmlee, S. Pellett, and R. A. Welch. *J. Bacteriol.* 163:94–105 (1985).
62. S. Fennewald, V. van Santen, and E. Kieff. *J. Virol.* 51:411–419 (1984).
63. S. Fields, G. Winter, and G. G. Brownlee. *Nature* 290:213–217 (1981).
64. P. J. Fleming, H. A. Dailey, D. Corcoran, and P. Strittmatter. *J. Biol. Chem.* 253:5369–5372 (1978).
65. R. M. Flugel, A. Rethwilm, B. Maurer, and G. Darai. *EMBO. J.* 6:2077–2084 (1987).
66. S. Froshauer and J. Beckwith. *J. Biol. Chem.* 259:10896–10903 (1984).
67. K. J. Fryxell and E. M. Meyerowitz. *EMBO. J.* 6:443–451 (1987).
68. W. J. Gallin, B. C. Sorkin, G. M. Edelman, and B. A. Cunningham. *Proc. Natl. Acad. Sci. USA* 84:2808–2812 (1987).
69. H. Garoff, A. Frischauf, K. Simons, H. Lehrach, and H. Delius. *Nature* 288:236–241 (1980).
70. D. P. Gold, J. M. Puck, C. L. Pettey, M. Cho, J. Coligan, J. N. Woody, and C. Terhorst. *Nature* 321:431–434 (1986).
71. M. B. Gorin, S. B. Yancey, J. Cline, J.-P. Revel, and J. Horwitz. *Cell* 39:49–59 (1984).
72. P. Gros, J. Croop, and D. Housman. *Cell* 47:371–380 (1986).
73. B. Guiard. *EMBO. J.* 4:3265–3272 (1985).
74. K. M. Hager, S. M. Mandala, J. W. Davenport, D. W. Speicher, E. J. Benz, Jr., and C. W. Slayman. *Proc. Natl. Acad. Sci. USA* 83:7693–7697 (1986).
75. A. Hampe, M. Gobet, C. J. Sherr, and F. Galibert. *Proc. Natl. Acad. Sci. USA* 81:85–89 (1984).
76. M. Haniu, T. Iyanagi, P. Miller, T. D. Lee, and J. E. Shively. *Biochemistry* 25:7906–7911 (1986).
77. J. J. Hemperly, B. A. Murray, G. M. Edelman, and B. A. Cunningham. *Proc. Natl. Acad. Sci.* 83:3037–3041 (1986).
78. I. Hermans-Borgmeyer, D. Zopf, R. Ryseck, B. Hovemann, H. Betz, and E. D. Gundelfinger. *EMBO J.* 5:1503–1508 (1986).
79. J. Hermolin, J. Gallant, and R. H. Fillingame. *J. Biol. Chem* 258:14550–14555 (1983).
80. J. E. Hesse, L. Wieczorek, K. Altendorf, A. S. Reicin, E. Dorus, and W. Epstein. *Proc. Natl. Acad. Sci. USA* 81:4746–4750 (1984).
81. W. Hoffman. *J. Biol. Chem.* 260:11831–11837 (1985).
82. E. C. Holland, J. O. Leung, and K. Drickamer. *Proc. Natl. Acad. Sci. USA* 81:7338–7342 (1984).
83. W. Hunziker, M. Spiess, G. Semenza, and H. F. Lodish. *Cell* 46:227–234 (1986).
84. K. Ikuta, M. Takami, C. Won Kim, T. Honjo, T. Miyoshi, Y. Tagaya, T. Kawabe, and J. Yodoi. *Proc. Natl. Acad. Sci. USA* 84:819–823 (1987).
85. D. Johnson, A. Lanahan, C. R. Buck, A. Sehgal, C. Morgan, E. Mercer, M. Bothwell, and M. Chao. *Cell* 47:545–554 (1986).
86. J. Kaput, S. Goltz, and G. Blobel. *J. Biol. Chem.* 257:15054–15058 (1982).
87. T. Kubo, K. Fukuda, A. Mikami, A. Maeda, H. Takahashi, M. Mishina, T. Haga, K. Haga, A. Ichiyama, K. Kangawa, M. Kojima, H. Matsuo, T. Hirose, and S. Numa. *Nature* 323:411–416 (1986).
88. R. A. Lamb, S. L. Zebedee, and C. D. Richardson. *Cell* 40:627–633 (1985).
89. Y. Laperche, F. Bulle, T. Aissani, M. Chobert, M. Aggerbeck, J. Hanoune, and G. Guellaen. *Proc. Natl. Acad. Sci. USA* 83:937–941 (1986).
90. R. A. Laursen, M. Samiullah, and M. B. Lees. *Proc. Natl. Acad. Sci. USA* 81:2912–2916 (1984).
91. C. A. Lee and M. H. Saier, Jr. *J. Biol. Chem.* 258:10761–10767 (1983).
92. D. C. Lee, T. M. Rose, N. R. Webb, and G. J. Todaro. *Nature* 313:489–491 (1985).
93. J. S. Lee, J. Trowsdale, P. J. Travers, J. Carey, F. Grosveld, J. Jenkins, and W. F. Bodmer. *Nature* 299:750–752 (1982).
94. D. R. Littman, M. Newton, D. Crommie, S. Ang, J. G. Seidman, S. N. Gettner, and A. Weiss. *Nature* 326:85–88 (1987).
95. D. H. MacLennan, C. J. Brandl, B. Korczak, and N. M. Green. *Nature (London)* 316:696–700 (1985).
96. P. I. MacKenzie. *J. Biol. Chem.* 262:9744–9749 (1987).
97. P. I. MacKenzie. *J. Biol. Chem.* 261:6119–6125 (1986).
98. D. Mathis, C. Benoist, V. Williams, M. Kanter, and H. McDevitt. *Cell* 32:745–754 (1988).
99. H. Matsushime, L. Wang, and M. Shibuya. *Mol. Cell. Biol.* 6:3000–3004 (1986).
100. W. Min Jou, M. Verhoeyen, R. Devos, E. Saman, R. Fang, D. Huylebroeck, and W. Fiers. *Cell* 19:683–696 (1980).
101. K. E. Mostov, M. Friedlander, and G. Blobel. *Nature* 308:37–43 (1984).
102. M. Mueckler, C. Caruso, S. A. Baldwin, M. Panico, I. Blench, H. R. Morris, W. J. Allard, G. E. Lienhard, and H. F. Lodish. *Science* 229:941–945 (1985).
103. Y. Nakashima, B. Frangione, R. L. Wiseman, and W. H. Konigsberg. *J. Biol. Chem.* 256:5792–5797 (1981).
104. J. Nathans and D. S. Hogness. *Cell* 34:807–814 (1983).

105. J. Nathans and D. S. Hogness. *Proc. Natl. Acad. Sci. USA* 81:4851–4855 (1984).
106. W. S. Neckameyer, M. Shibuya, M. Hsu, and L. Wang. *Mol. Cell. Biol.* 6:1478–1486 (1986).
107. W. S. Neckameyer and L. Wang. *J. Virol.* 53:879–884 (1985).
108. M. Noda, H. Takahashi, T. Tanabe, M. Toyosato, S. Kikyotani, Y. Furutani, T. Hirose, H. Takashima, S. Inayama, T. Miyata, and S. Numa. *Nature* 302:528–532 (1983).
109. S. Paabo, F. Weber, T. Nilsson, W. Schaffner, and P. A. Peterson. *EMBO J.* 5:1921–1927 (1986).
110. E. G. Peralta, J. W. Winslow, G. L. Peterson, D. H. Smith, A. Ashkenazi, J. Ramachandran, M. I. Schimerlik, and D. J. Capon. *Science* 236:600–605 (1987).
111. S. J. Ralph, M. L. Thomas, C. C. Morton, and I. S. Trowbridge. *EMBO J.* 6:1251–1257 (1987).
112. J. V. Ravetch, A. D. Luster, R. Weinshank, J. Kochan, A. Pavlovec, D. A. Portnoy, J. Hulmes, Y. E. Pan, and J. C. Unkeless. *Science* 234:718–725 (1986).
113. N. K. Robakis, N. Ramakrishna, G. Wolfe, and H. M. Wisniewski. *Proc. Natl. Acad. Sci. USA* 84:4190–4194 (1987).
114. A. J. M. Roebroek, J. A. Schalken, J. A. M. Leunissen, C. Onnekink, H. P. J. Bloemers, and W. J. M. Van de Ven. *EMBO J.* 5:2197–2200 (1986).
115. J. Rogers, P. Early, C. Carter, K. Calame, M. Bond, L. Hood, and R. Wall. *Cell* 20:303–312 (1980).
116. P. J. M. Rottier, G. W. Welling, S. Welling-Wester, H. G. M. Neisters, J. A. Lenstra, and B. A. M. Van der Zeijst. *Biochemistry* 25:1335–1339 (1986).
117. J. K. Rose and J. E. Bergmann. *Cell* 30:753–762 (1982).
118. D. W. Russell, W. J. Schneider, T. Yamamoto, K. L. Luskey, M. S. Brown, and J. L. Goldstein. *Cell* 37:577–585 (1984).
119. A. F. Russo and D. E. Koshland, Jr. *Science* 220:1016–1020 (1983).
120. Y. Saga, J. Tung, F. Shen, and E. A. Boyse. *Proc. Natl. Acad. Sci. USA* 83:6940–6944 (1986).
121. H. Saito, D. M. Kranz, Y. Takagaki, A. C. Hayday, H. N. Eisen, and S. Tonegawa. *Nature* 309:757–762 (1984).
122. L. Schenning, D. Larhammar, P. Bill, K. Wiman, A. Jonsson, L. Rask, and P. A. Peterson. *EMBO J.* 3:447–452 (1984).
123. C. Schneider, M. J. Owen, D. Banville, and J. G. Williams. *Nature (London)* 311:675–678 (1984).
124. T. Seki, T. Moriuchi, H. Chang, R. Denome, and J. Silver. *Nature* 313:485–487 (1985).
125. W. A. Sewell, M. H. Brown, J. Dunne, M. J. Owen, and M. J. Crumpton. *Proc. Natl. Acad. Sci. USA* 83:8718–8722 (1986).
126. G. E. Shull, L. K. Lane, and J. B. Lingrel. *Nature* 321:429–431 (1986).
127. G. E. Shull, and J. B. Lingrel. *J. Biol. Chem.* 261:16788–16791 (1986).
128. G. E. Shull, A. Schwartz, and J. B. Lingrel. *Nature (London)* 316:691–695 (1985).
129. M. Spiess, A. L. Schwartz, and H. F. Lodish. *J. Biol. Chem.* 260:1979–1982 (1985).
130. M. Strubin, B. Mach, and E. O. Long. *EMBO J.* 3:869–872 (1984).
131. S. Suzuki, W. S. Argraves, R. Pytela, H. Arai, T. Krusius, M. D. Pierschbacher, and E. Ruoslahti. *Proc. Natl. Acad. Sci.* 83:8614–8618 (1986).
132. J. W. Tamkun, D. W. DeSimone, D. Fonda, R. S. Patel, C. Buck, A. F. Horwitcz, and R. O. Hynes. *Cell* 46:271–282 (1986).
133. G. E. Tarr, S. D. Black, V. S. Fujita, and M. J. Coon. *Proc. Natl. Acad. Sci. USA* 80:6552–6556 (1983).
134. M. Tomita and V. T. Marchesi. *Proc. Natl. Acad Sci USA* 72:2964–2968 (1975).
135. J. Trowsdale and A. Kelly. *EMBO J.* 4:2231–2237 (1985).
136. P. van den Elsen, B. Shepley, J. Borst, J. E. Coligen, A. F. Markham, S. Orkin, and C. Terhorst. *Nature* 312:413–418 (1984).
137. P. van den Elsen, B. Shepley, M. Cho, and C. Terhorst. *Nature* 314:542–545 (1985).
138. M. Verhoeyen, R. Fang, W. M. Jou, R. Devos, D. Huylebroeck, E. Samam, and W. Fiers. *Nature* 286:771–776 (1980).
139. H. Vogel, J. K. Wright, and F. Jahnig. *EMBO J.* 4:3625–3631 (1985).
140. J. E. Walker, M. Sarastre, and N. J. Gasy. *Biochim. Biophys. Acta* 768:164–200 (1984).
141. R. J. Watson, J. H. Weis, J. S. Salstrom, and L. W. Enquist. *Science* 218:381–384 (1982).
142. W. R. Widger, W. A. Cramer, R. G. Herrmann, and A. Trebst. *Proc. Natl. Acad. Sci. USA* 81:674–678 (1984).
143. W. S. M. Wold, C. Cladaras, S. L. Deutscher, and Q. S. Kapoor. *J. Biol. Chem.* 260:2424–2431 (1985).
144. Y. Yamawaki-Kataoka, S. Nakai, T. Miyata, and T. Honjo. *Proc. Natl. Acad. Sci. USA* 79:2723–2627 (1982).
145. T. Yamamoto, S. Ikawa, T. Akiyama, K. Semba, N. Nomura, N. Miyajima, T. Saito, and K. Toyoshima. *Nature* 319:230–234 (1986).
146. T. Yamamoto, C. G. Davis, M. S. Brown, W. J. Schneider, M. L. Casey, J. L. Goldstein, and D. W. Russell. *Cell* 39:27–38 (1984).
147. H. Yazyu, S. Shiota-Niiya, T. Shimamoto, H. Kanazawa, Futai M., and T. Tsuchiya. *J. Biol. Chem.* 259:4320–4326 (1984).
148. Y. Yarden, H. Rodriguez, S. K. F. Wong, D. R. Brandt, D. C. May, J. Burnier, K. N. Harkins, E. Y. Chen, J. Ramachandran, A. Ullrich, and E. M. Ross. *Proc. Natl. Acad. Sci.* USA 83:6795–6799 (1986).
149. R. Serrano, M. C. Kielland-Brandt, and G. R. Fink. *Nature* 319:689–693 (1986).
150. D. Larhammar, L. Schenning, K. Gustafsson, K. Wiman, L. Claesson, L. Rask, and P. A. Peterson. *Proc. Natl. Acad. Sci.* 79:3687–3691 (1982).
151. P. B. Wolfe, W. Wickner, and J. M. Goodman. *J. Biol. Chem.* 258:12073–12080 (1983).
152. J. H. Elder and J. I. Mullins. *J. Virol.* 46:871–880 (1982).
153. S. Fields, G. Winter, and G. G. Brownler. *Nature* 270:213–218 (1981).