Annu. Rev. Biophys. Biophys. Chem. 1990. 19:369–403

ON THE MICROASSEMBLY OF INTEGRAL MEMBRANE PROTEINS[1]

Jean-Luc Popot and Catherine de Vitry

Service de Photosynthèse, Institut de Biologie Physico-Chimique,
13 rue Pierre-et-Marie Curie, F-75005 Paris, France

KEY WORDS: protein traffic, folding domains, organelle membrane proteins, integral membrane protein subunits, protein import into chloroplast and mitochondrion.

CONTENTS

PERSPECTIVES AND OVERVIEW 370
PROCEDURES 371
Hydrophobicity Analysis 371
Choice of Proteins 373
RESULTS 374
Identifying Putative Transmembrane Segments and Estimating Their Hydrophobicity 374
Number and Hydrophobicity of Putative Transmembrane α-Helices as a Function of Protein Localization and Function 382
Number and Hydrophobicity of Transmembrane Segments in Organelle Inner-Membrane Proteins Depending on Site of Synthesis 383
DISCUSSION 387
The Microassembly of Integral Membrane Proteins in Organelles 387
Other Membranes 395
Displacing the Synthesis of Integral Membrane Proteins from Organelles to the Cytoplasm 397
CONCLUSION 399

[1] This review is dedicated to Annemarie Weber (University of Pennsylvania) and Andrew G. Szent-Györgyi (Brandeis University), instructors at the Physiology Summer Course at Woods Hole, Massachusetts, in 1971.

0883–9182/90/0610–0369$02.00

PERSPECTIVES AND OVERVIEW

Recent experimental evidence suggests that the folding of transmembrane regions of integral membrane proteins should be regarded as qualitatively different from that of soluble proteins. In soluble proteins, hierarchical levels of folding consist of the secondary structure units (α-helices, β-strands), super-secondary motifs (e.g. the $\beta\alpha\beta$ unit or the 8-fold $\alpha\beta$ barrel), domains (comprised of one or several motifs), protomers (comprised of one or several domains), and, if a quaternary structure exists, oligomers (cf 44, 113, 123). Folding is believed to follow this sequence more or less closely, with the first independently stable structures appearing at the domain level (e.g. 39, 64, 144). Domains vary in size. For globular proteins, they comprise typically 70–150 amino acid residues. Smaller domains, e.g. the 40–50 residue domains in rubredoxin or wheat germ agglutinin, are found in proteins stabilized by disulfide bridges or prosthetic groups (for reviews, see 65, 113, 123, 144). As a consequence, soluble proteins or protein subunits smaller than ~ 70 residues are very rare (129a).

In the transmembrane regions of integral membrane proteins, a single hydrophobic α-helix apparently has, to a large extent, the properties of a domain in itself. Both theoretical considerations and experimental observations form the basis for this view (for a review, see 103). First, transmembrane regions in bacterial photosynthetic reaction centers and in bacteriorhodopsin are composed of hydrophobic α-helices. Free energy estimates lead to the prediction that each α-helix can form an independent, stable, transmembrane entity in a lipid bilayer. Second, several integral membrane proteins have been functionally reassembled starting from fragments that had been independently refolded or synthesized (Table 1). In these experiments, each fragment folded autonomously, as expected if it is itself comprised of elements (α-helices) that can take up a largely correct transmembrane position and secondary structure by themselves. In addition, two natural cases suggest such a process, in which a single polypeptide in one membrane or organism appears in another to be split into two subunits (Table 1).

An α-helix long enough to cross the ~ 30-Å thick fatty acyl region of a phospholipid bilayer comprises 20 residues. Transmembrane regions may thus be built up of stable units that are considerably smaller than those involved in the folding of soluble proteins. The possibility arises that this particularity dictates some characteristic features of membrane protein biosynthesis and assembly.

In the present review, we survey the subunit composition, size, and number of hydrophobic transmembrane segments of most eukaryotic integral membrane proteins for which the sequence is known (see also 32a).

Table 1 Integral membrane proteins assembled from fragments that had been either refolded or biosynthesized independently

Origin of fragments	Medium where assembly occurs	Protein	Hydrophobic helices per fragment	References
Proteolysis	lipid/detergent mixed micelles	bacteriorhodopsin	2+5	62, 78
Proteolysis	lipid/detergent mixed micelles	bacteriorhodopsin	5+2	77, 127
Proteolysis	lipid vesicles	bacteriorhodopsin	2+5	104, 105
Proteolysis	lipid vesicles	bacteriorhodopsin	5+2	68
Proteolysis	lipid vesicles	bacteriorhodopsin	1+1+5	69
Engineered mRNA	*Xenopus* oocyte (ER?)	β_2 adrenergic receptor	5+2	71
Engineered plasmids	*E. coli* plasma membrane	*lac* permease	2+12	147
Natural[a]	thylakoid	cytochrome b_6 +subunit IV	4+3	57, 145a
Natural[a]	*E. coli* plasma membrane	Nicotinamide nucleotide transhydrogenase	8+4	148

[a] Natural cases correspond to proteins composed of one polypeptide chain in one type of membrane or organism and two distinct subunits in another. The number of putative hydrophobic transmembrane segments in cytochrome b_6, subunit IV, and nicotinamide nucleotide transhydrogenase is discussed in the footnotes to Table 4. Cytochrome b_6 and subunit IV from chloroplast b_6/f complex are respectively homologous to the amino terminal and carboxyterminal parts of cytochrome b from the bc_1 complexes from mitochondria or purple bacteria; a segment homologous to the last of the 8 putative hydrophobic transmembrane α-helices in cytochrome b is missing in subunit IV. The α and β subunits of *E. coli* transhydrogenase are homologous to the aminoterminal and carboxyterminal parts of the beef enzyme, respectively.

This analysis shows that many of them indeed contain subunits that are much smaller than subunits of soluble proteins. Very marked differences in composition and properties exist depending on which membrane the proteins lie in. A particularly striking observation is that the composition of complexes from the inner mitochondrial membrane and thylakoid membrane apparently reflects a restriction on the import of large hydrophobic proteins from the cytoplasm, while including a large number of very small, 1- or 3-α-helix subunits.

PROCEDURES

Hydrophobicity Analysis

The sequences of about 250 presumed or proven integral membrane proteins were taken from the CITI2 (Paris) data base BISANCE (data banks

NBRF, EMBL, GENPRO, and GENBANK) or collected from the literature. We examined them for the presence of potential transmembrane hydrophobic α-helices using a modification of Klein et al's (70) program run on a Vax 750 computer. The program examined each sequence twice. In the first pass, using the hydrophobicity scale of Kyte & Doolittle [KD scale (73)] and a 17-residue span, the program generated the number and approximate limits of the putative transmembrane helices, together with an index of the relative probability (P/I), that each segment is either peripheral or transmembrane (70). The program also gave the average hydrophobicity of each segment (GES) expressed in kcal/residue using the hydrophobicity scale of Engelman et al [Table 2, GES scale (37)]. In the second pass, the search was done using the GES scale and a 17-residue span. The procedure is similar to that applied by von Heijne (142a) to bacterial membrane protein sequences.

Both searches usually identified the same hydrophobic segments. In some cases, neighboring hydrophobic segments separated by a few polar residues were properly distinguished in the first pass and not in the second, or vice versa. Topological models from the literature were then compared to results from the hydrophobicity analyses as described in the section covering results and in footnotes to Table 4. For the purpose of estimating helix hydrophobicities, the limits of the segments were defined using the GES scale except for a dozen cases (over about 600 segments) in which segments matching those proposed in the literature were better identified using the KD scale. In all cases and throughout this text, hydrophobicities are expressed as GES values.

To test the reliability of our approach in estimating the hydrophobicity of putative transmembrane helices, we applied it to the three integral

Table 2 Goldman-Engelman-Steitz (GES) hydrophobicity scale

Residue	Transfer free energy[a]	Residue	Transfer free energy	Residue	Transfer free energy
Phenylalanine	3.7	Alanine	1.6	Glutamine	−4.1
Methionine	3.4	Threonine	1.2	Asparagine	−4.8
Isoleucine	3.1	Glycine	1.0	Glutamic acid	−8.2
Leucine	2.8	Serine	0.6	Lysine	−8.8
Valine	2.6	Proline	−0.2	Aspartic acid	−9.2
Cysteine	2.0	Tyrosine	−0.7	Arginine	−12.3
Tryptophan	1.9	Histidine	−3.0		

[a] Free energy (kcal/residue) for transferring residues in an α-helix from a nonaqueous environment to water (from Ref. 37).

membrane proteins in the photosynthetic reaction center from purple bacteria. In all cases, the 17-residue segments identified as the most hydrophobic using either the KD or the GES scales corresponded to the transmembrane helices. In 10 cases (out of 11 helices), the segments were included within the transmembrane helices of the electron density map (30a) to within 3 residues. The GES scale displaced helix E in subunit L by 7 residues past the end of the transmembrane region, while the KD scale placed it correctly to within 1 residue; the difference in GES value depending on which segment limits were used was 0.09 kcal/residue. Conversely, using the KD scale resulted in misplacing helix C in subunit M by 6 residues, while the GES scale placed it correctly. The hydrophobicity difference was 0.04. Clearly, as expected, inaccuracy in defining helix limits entailed only small errors on GES.

The transmembrane segments in the photoreaction center are fairly hydrophobic, which simplifies identification. In bacteriorhodopsin, the 17-residue segments identified as the most hydrophobic using either scale again matched accepted transmembrane helices (37). The least hydrophobic helix (helix C, which contains two aspartic acid residues), was predicted at the same position (within 1 residue) by the two scales. The greatest discrepancy between the hydrophobicity estimates calculated using the limits given by the two scales reached 0.13 (helix G). The results obtained on bacteriorhodopsin and the reaction center proteins are summarized in Table 3.

Choice of Proteins

We distributed integral membrane proteins into four sets:

1. proteins from the plasma membrane of eukaryotic cells and of other

Table 3 Number and hydrophobicity of transmembrane helices in four well-characterized integral proteins from bacterial cytoplasmic membranes

	GES of transmembrane segments[a]							
	Segment position in sequence							
Protein	A	B	C	D	E	F	G	GES_{av}[b]
Reaction center (*R. viridis*):								
Subunit H	1.63							1.63
Subunit L	2.60	1.86	2.11	1.70	1.89			2.03
Subunit M	2.31	2.15	2.02	1.71	1.86			2.01
Bacteriorhodopsin	2.02	2.01	0.94	1.50	2.13	1.74	1.25	1.66

[a] Average free energy of transfer for the most hydrophobic 17-residue segment overlapping each helix (kcal/residue).
[b] Average of the individual GES values.

membranes that are directly in contact with the cytosol (endoplasmic and sarcoplasmic reticulum, retina sacculae, exocytotic vesicles), which we designate as *plasma membrane proteins*. Several homologous proteins from the same or different species were often analyzed, although the results for only one of them are incorporated in the Figures and analysis.

2. proteins from the mitochondrial inner membrane, whether coded for by mitochondrial or nuclear DNA. Mitochondrial genomes were analyzed in totality for *Bos primegenius taurus*, *Homo sapiens*, *Mus musculus*, and *Xenopus laevis*, and partially, as a function of available sequences, for *Neurospora crassa*, *Saccharomyces cerevisiae*, *Leishmania tarentolae*, *Trypanosoma brucei*, *Chlamydomonas reinhardtii*, and *Zea mays*. We also analyzed the proteins of the mitochondrial inner membrane of these species encoded by nuclear DNA and of known sequence. The results are shown for bovine proteins except if the sequence was only available for other eukaryotes. The two hydrophobic subunits of the succinate dehydrogenase complex have similar molecular weights in eukaryotes and in the prokaryote *Escherichia coli* and, to our knowledge, have not been sequenced in any eukaryotes. In this particular case, we show the results for *E. coli*.

3. proteins from the thylakoid membrane, whether encoded in the chloroplast or the nucleus, and chloroplast DNA open reading frames (ORFs) encoding putative, unidentified proteins. The chloroplast genomes of *Marchantia polymorpha* and *Nicotiana tabacum* were analyzed in totality, although only the results for *Marchantia* are shown, except when the subunits were better characterized in *Nicotiana*. The integral proteins of the thylakoid membranes encoded in the nucleus were analyzed in several higher plants as well as in *C. reinhardtii*, as a function of available sequences.

4. proteins from the outer membrane of *E. coli*.

When several homologous proteins from different organisms or representing related enzymes were examined, only one protein of each family was included in the final analysis unless sequence similarities were less than 35–40%. The alignment program (alignment score program of B. C. Orcutt, M. O. Dayhoff, and W. C. Barker, 1984 version available at the CITI2) was based on the algorithm of Needleman & Wunsch (89) using a unitary matrix and a break penalty parameter of 8.

RESULTS

Identifying Putative Transmembrane Segments and Estimating Their Hydrophobicity

Table 4 lists 140 integral membrane proteins, which covers nearly all eukaryotic integral membrane proteins of known sequence when proteins

Table 4 Number and hydrophobicity of putative hydrophobic transmembrane segments in integral proteins from eukaryotic membranes[a]

4A Proteins from membranes that are directly in contact with the cytosol (plasma membrane, endoplasmic and sarcoplasmic reticulum, retina sacculae, exocytotic vesicles)[b]

Protein	*Genus*	N_r	N_h	*G E S* min	*G E S* av.	*G E S* max.	*Ref.*
HLA cl.II DN. A (DZα) gene product	*Homo*	225	1		2.42		NBRF
H2 class I "37" protein	*Mus*	337	1		2.49		73a
T-cell L3T4 glycoprotein	*Mus*	435	1		2.32		NBRF
T-cell CD3 glycoprotein δ chain	*Homo*	150	1		1.99		NBRF
T-cell receptor β-chain precursor	*Oryctolagus*	(319)	1		2.11		NBRF
B-lymphocyte glycoprotein PC_1	*Homo*	905	1		2.23		NBRF
Uvomorulin	*Mus*	728	1		2.62		114
Asialoglycoprotein receptor	*Rattus*	284	1		2.42		NBRF
Fibronectin receptor α chain	*Homo*	1008	1		2.66		EMBL
Fibronectin receptor β chain	*Gallus*	803	1		2.31		134
N-CAM	*Gallus*	1072	1		2.62		29
Glycophorin A	*Homo*	131	1		2.47		NBRF
Glycophorin C#	*Homo*	128	1		2.52		NBRF
ILGF-II receptor	*Homo*	2451	1		2.33		84
PDGF receptor	*Mus*	1067	1		2.59		149
EGF receptor	*Homo*	1186	1		2.42		NBRF
EGF receptor	*Drosophila*	1330	1		2.37		NBRF
NGF receptor	*Homo*	399	1		2.07		NBRF
Lymphocyte IgE receptor	*Homo*	321	1		2.18		NBRF
Interleukin-2 receptor P55 chain	*Homo*	251	1		2.27		NBRF
Insulin receptor	*Homo*	(1370)	1		2.49		NBRF
Transferrin receptor	*Homo*	760	1		2.46		NBRF
LDL receptor	*Homo*	860	1		2.57		NBRF
Toll gene product	*Drosophila*	(1097)	1		2.33		54
lin-12 gene product	*Caenorhabditis*	1429	1		2.54		150
Notch gene product	*Drosophila*	(2703)	1		2.56		NBRF
Thy-1 antigen	*Homo*	142	1		1.69		NBRF
CSF-1 receptor (c-*fms*)	*Homo*	(959)	1		2.57		25
High affinity IgE rec. α chain	*Rattus*	222	1		1.57		10
High affinity IgE rec. β chain#	*Rattus*	243	4	1.75	2.18	2.55	10
High affinity IgE rec. γ chain#	*Rattus*	62	1		1.51		10
β-gal. α 2,6-sialyltransferase	*Rattus*	403	1		2.71		Gpro
Aminopeptidase N	*Homo*	967	1		2.28		94
Enkephalinase	*Rattus*	750	1		2.39		Gpro
Guanylate cyclase	*Arbacia*	955	1		2.01		129
Intestinal sucrase-isomaltase	*Oryctolagus*	(1827)	1		2.77		Gbk
Cytochrome b_5	*Bos*	135	1		1.78		NBRF
Stearyl-CoA desaturase	*Rattus*	358	(4)	1.32	1.78	2.04	136
Cytochrome P-450 (C21)	*Bos*	496	(1)		1.91		NBRF
NADPH-cyt. P-450 oxidoreductase	*Rattus*	678	1		2.41		NBRF
HMG-CoA reductase	*Cricetulus*	887	7	0.68	1.58	2.29	NBRF
Synaptophysin	*Rattus*	307	4	1.70	1.97	2.27	75
Myelin proteolipid	*Bos*	276	(4)	1.99	2.17	2.29	NBRF
Opsin	*Bos*	348	7	1.41	2.00	2.38	NBRF
Opsin	*Drosophila*	373	7	0.70	1.69	2.06	NBRF
M_2 muscarinic receptor	*Homo*	466	7	0.63	1.75	2.36	99
β_2 adrenergic receptor	*Cricetulus*	418	7	0.79	1.83	2.54	NBRF
D_2 dopaminergic receptor	*Rattus*	415	7	1.13	1.90	2.57	16

continued

Table 4 (*continued*)

Protein	*Genus*	N_r	N_h	*GES* min	*GES* av.	*GES* max.	*Ref.*
1_c serotonin receptor	*Rattus*	460	7	1.33	2.04	2.75	66
α-factor receptor (STE2)	*Saccharomyces*	431	7	0.99	1.69	2.13	NBRF
a-factor receptor (STE3)	*Saccharomyces*	470	7	0.92	2.05	2.74	NBRF
Substance K receptor	*Bos*	384	7	1.01	1.85	2.49	Gpro
mas oncogene	*Homo*	325	7	1.22	1.97	2.75	NBRF
LH-hCG receptor[$]	*Sus*	674	7	1.17	1.75	2.24	79
Ca^{2+} ATPase (slow twitch muscle)	*Oryctolagus*	997	10	0.91	1.55	1.99	NBRF
Ca^{2+} ATPase (plasma membrane)	*Homo*	1220	10	0.45	1.74	2.42	140
Na^+/K^+ ATPase α chain	*Ovis*	1021	(7)	1.40	1.97	2.42	NBRF
Na^+/K^+ ATPase β chain	*Gallus*	305	1		2.28		Gbk
Na^+/K^+ ATPase putative γ chain[#]	*Ovis*	68	1		2.53		24
H^+ ATPase (plasma membrane)	*Neurospora*	920	(10)	1.08	1.83	2.57	Gpro
H^+/K^+ ATPase	*Rattus*	1016	7	1.15	1.96	2.44	Gpro
Adenylyl cyclase[$]	*Bos*	1134	12	1.34	1.85	2.66	72
Uracil transport protein	*Saccharomyces*	633	12	1.19	1.79	2.56	67
Glucose transporter	*Homo*	492	12	1.00	1.89	2.75	Gbk
Na^+/glucose co-transporter	*Oryctolagus*	662	(15)	1.35	1.95	2.58	56
Arginine permease	*Saccharomyces*	590	7	1.48	1.98	2.37	Gbk
P-glycoprotein (*mdr*1)	*Homo*	1280	12	1.22	1.90	2.59	NBRF
patched gene product[$]	*Drosophila*	1299	12	1.41	2.01	2.43	88
Anion exchange protein	*Mus*	929	13	1.05	1.91	2.59	Gbk
Nicotinic ACh receptor α chain	*Torpedo*	437	4	2.22	2.44	2.64	NBRF
Glycine receptor 48K chain	*Rattus*	421	4	1.47	1.88	2.09	47
$GABA_A$ receptor α chain	*Bos*	429	4	1.66	1.78	1.97	122
Voltage gated Na^+ channel	*Electrophorus*	1820	(20)	0.12	1.79	2.81	NBRF
Ca^{2+} channel α_1 subunit	*Oryctolagus*	1873	(20)	0.56	1.80	2.88	135
K^+ channel (*Shaker* gene)	*Drosophila*	616	(5)	1.39	2.03	2.61	102
Ryanodine receptor	*Oryctolagus*	5037	4	1.99	2.20	2.49	133
Inositol trisphosphate receptor[$]	*Mus*	2749	7	1.00	1.86	2.55	41a
Lens fiber MP26	*Bos*	263	6	1.59	1.71	1.89	Gbk
Gap junction connexin	*Rattus*	283	4	1.69	1.89	2.36	Gbk

4B Proteins from the inner mitochondrial membrane[c]

Mitochondrion-encoded proteins	*Genus*	N_r	N_h	*GES* min	*GES* av.	*GES* max.	*Ref.*
NADH-Q reductase, subunit 1	*Bos*	318	(8)	1.37	2.05	2.54	NBRF
NADH-Q reductase, subunit 2	*Bos*	347	10	1.55	2.03	2.74	NBRF
NADH-Q reductase, subunit 3	*Bos*	115	3	1.95	2.17	2.51	NBRF
NADH-Q reductase, subunit 4	*Bos*	459	15	1.34	1.81	2.36	NBRF
NADH-Q reductase, subunit 5	*Bos*	606	15	0.96	1.91	2.51	NBRF
NADH-Q reductase, subunit 6	*Bos*	175	5	1.88	2.25	2.64	NBRF
NADH-Q reductase, subunit 4L	*Bos*	98	3	1.66	1.84	1.93	NBRF
QH2-cyt. *c* reductase, cytochrome *b*	*Bos*	379	(8)	1.26	2.11	2.66	NBRF
Cyt. *c* oxidase, subunit I	*Bos*	514	12	1.45	2.10	2.72	NBRF
Cyt. *c* oxidase, subunit II	*Bos*	227	2	2.17	2.27	2.38	NBRF
Cyt. *c* oxidase, subunit III	*Bos*	261	7	1.37	1.76	2.11	NBRF
ATPase, subunit 6	*Bos*	226	(5)	1.67	1.94	2.22	NBRF
ATPase, subunit 8[$]	*Bos*	66	1		2.62		NBRF
Nucleus-encoded proteins							
Succinate-Q reductase, subunit C	*(Escherichia)*	128	(3)	1.84	2.13	2.30	146
Succinate-Q reductase, subunit D	*(Escherichia)*	115	(3)	2.12	2.15	2.30	146

Table 4 (*continued*)

Nucleus-encoded proteins	*Genus*	N_r	N_h	*G E S* min	av.	max.	*Ref.*
QH2-cyt. *c* reductase, cytochrome c_1	*Bos*	241	1		1.69		NBRF
QH2-cyt. *c* reductase, FeS protein°	*Bos*	196	1		1.51		120
QH2-cyt. *c* reductase, subunit VII#	*Bos*	81	1		1.22		11
QH2-cyt. *c* reductase, subunit X#	*Bos*	62	1		0.96		NBRF
QH2-cyt. *c* reductase, subunit XI#	*Bos*	56	1		1.44		119
Cytochrome *c* oxidase, subunit IV	*Bos*	147	1		2.49		NBRF
Cytochrome *c* oxidase, subunit VIIIa#	*Bos*	47	1		2.39		NBRF
Cytochrome *c* oxidase, subunit VIIIb#	*Bos*	46	1		2.04		NBRF
ATPase, subunit 9	*Bos*	75	2	1.86	2.28	2.71	NBRF
Threonine deshydratase°	*Saccharomyces*	576	1		1.92		NBRF
Nicotinamide nucl. transhydrogenase°	*Bos*	1043	(12)	1.51	1.74	1.97	148
ADP/ATP carrier protein°	*Bos*	297	(3)	1.13	1.45	1.76	NBRF
Brown fat uncoupling protein°	*Cricetulus*	306	3	1.41	1.61	1.96	NBRF
Phosphate carrier protein°	*Bos*	313	3	0.68	1.53	2.12	NBRF

4C Proteins from the thylakoid membrane[d]

Chloroplast-encoded proteins	*Genus*	N_r	N_h	*G E S* min	av.	max.	*Ref.*
NADH-Q reductase, subunit 1	*Marchantia*	368	(7)	1.54	1.93	2.63	NBRF
NADH-Q reductase, subunit 2	*Marchantia*	501	15	1.28	1.90	2.56	NBRF
NADH-Q reductase, subunit 3	*Marchantia*	120	3	1.38	2.07	2.79	NBRF
NADH-Q reductase, subunit 4	*Marchantia*	499	14	1.27	1.85	2.42	NBRF
NADH-Q reductase, subunit 5	*Marchantia*	692	17	1.13	2.04	2.69	NBRF
NADH-Q reductase, subunit 6	*Marchantia*	191	5	1.61	2.03	2.46	NBRF
NADH-Q reductase, subunit 4L	*Marchantia*	100	3	0.91	1.68	2.45	NBRF
PSII, subunit D1 (*psbA*)	*Marchantia*	343	5	1.15	1.82	2.27	NBRF
PSII, subunit 47kDa (*psbB*)	*Marchantia*	508	6	1.48	1.97	2.24	NBRF
PSII, subunit 44kDa (*psbC*)	*Marchantia*	459	6	1.60	1.97	2.45	NBRF
PSII, subunit D2 (*psbD*)	*Marchantia*	353	5	1.91	2.11	2.29	NBRF
PSII, cytochrome b_{559} (*psbE*)#	*Nicotiana*	83	1		1.99		Gpro
PSII, cytochrome b_{559} (*psbF*)#	*Nicotiana*	39	1		1.82		Gpro
PSII, phosphosubunit (*psbH*)#	*Nicotiana*	73	1		2.54		Gpro
PSII, subunit encoded by *psbI*#	*Marchantia*	36	1		2.54		NBRF
PSII, subunit encoded by *psbJ*#	*Marchantia*	40	1		2.35		NBRF
PSII, subunit encoded by *psbK*#	*Nicotiana*	37	1		2.18		86
PSII, subunit encoded by *psbL*#	*Marchantia*	38	1		2.24		NBRF
Cytochrome b_6f, cytochrome *f* (*petA*)	*Marchantia*	285	1		2.26		NBRF
Cytochrome b_6f, cytochrome b_6 (*petB*)	*Marchantia*	215	4	1.39	1.94	2.39	NBRF
Cytochrome b_6f, subunit IV (*petD*)	*Marchantia*	160	3	2.11	2.13	2.19	NBRF
PSI, subunit P700 (*psaA1*)	*Marchantia*	750	11	1.13	1.71	2.35	NBRF
ATPase, subunit I (*atpF*)	*Marchantia*	184	1		0.94		NBRF
ATPase, subunit III (*atpH*)	*Marchantia*	81	2	1.74	1.83	1.92	NBRF
ATPase, subunit IV (*atpI*)	*Marchantia*	248	(5)	1.71	2.01	2.30	NBRF
Nucleus-encoded proteins							
Cytochrome b_6f, Rieske FeS protein°	*Spinacia*	179	1		1.77		131
PSI, subunit P37#	*Chlamydomonas*	87	1		1.74		41
LHCII chlo. ab protein type I°	*Pisum*	233	3	1.22	1.66	2.39	17
LHCI chlo. ab protein type I°	*Lycopersicon*	202	3	1.19	1.39	1.62	61
LHCI chlo. ab protein type II°	*Petunia*	(179)	3	0.94	1.10	1.41	130
LHCI chlo. ab protein type III°	*Lycopersicon*	(241)	3	1.13	1.41	1.78	101

continued

4D Protein from the inner envelope membrane of chloroplast[d]

Protein	*Genus*	N_r	N_h	*GES min*	*GES av.*	*GES max.*	*Ref.*
Phosphate translocator°$	*Spinacia*	(404)	7	1.24	1.75	2.05	40

4E Hypothetical membrane proteins encoded by chloroplast open reading frames (ORFs)[d]

Protein	*Genus*	N_r	N_h	*GES min*	*GES av.*	*GES max.*	*Ref.*
Hypothetical protein 135	*Marchantia*	135	3	2.02	2.13	2.22	NBRF
Hypothetical protein 184	*Marchantia*	184	2	2.17	2.42	2.67	NBRF
Hypothetical protein 1068	*Marchantia*	1068	6	1.36	1.84	2.11	NBRF
Hypothetical protein 203	*Marchantia*	203	1		2.01		NBRF
Hypothetical protein 288	*Marchantia*	288	6	1.83	2.08	2.59	NBRF
Hypothetical protein 2136	*Marchantia*	2136	2	1.78	1.94	2.10	NBRF
Hypothetical protein 320	*Marchantia*	320	6	1.32	1.88	2.24	NBRF
Hypothetical protein 36b#	*Marchantia*	36	1		2.16		NBRF
Hypothetical protein 434	*Marchantia*	434	5	1.32	1.66	1.91	NBRF
Hypothetical protein 42b#	*Marchantia*	42	1		2.17		NBRF
Hypothetical protein 31#	*Marchantia*	31	1		2.54		NBRF
Hypothetical protein 32#	*Marchantia*	32	1		2.09		NBRF
Hypothetical protein 33#	*Marchantia*	33	1		2.11		NBRF
Hypothetical protein 34#	*Marchantia*	34	1		2.45		NBRF
Hypothetical protein 35#	*Marchantia*	35	1		2.64		NBRF
Hypothetical protein 37#	*Marchantia*	37	1		2.24		NBRF
Hypothetical protein 50#	*Marchantia*	50	1		2.16		NBRF
Hypothetical protein 55#	*Marchantia*	55	1		2.18		NBRF
Hypothetical protein 62	*Marchantia*	62	2	2.21	2.35	2.48	NBRF
Hypothetical protein mbpX	*Marchantia*	370	2	1.22	1.29	1.36	NBRF

[a] N_r is the number of residues in the mature protein, except in a few cases as indicated by parentheses. These cases are discussed in Footnotes b–d. N_h is the number of hydrophobic transmembrane segments in the currently accepted topological model for the protein. When no such model exists or for special cases, the value of N_h is placed between parentheses and discussed in Footnotes b–d. To limit the number of references, we have not systematically referred to original proposals for transmembrane topologies when they appear generally accepted and present no particular problem. These are usually accessible from the data banks or the references indicated. *GES av.* is the average hydrophobicity of the putative transmembrane segments, determined as described in the section on procedures and expressed in kcal/residue; *GES min.* is the hydrophobicity of the least hydrophobic of these segments, *GES max.* that of the most hydrophobic one. Sequences were taken from the references indicated or from the following data banks: EMBL, GENBANK (Gbk), GENPRO (Gpro), and NBRF. Abbreviations used in the names of proteins: *cl.*, class; *nucl.*, nucleotide; *β-gal.*, β-galactoside; *cyt.*, cytochrome; *msc.*, muscle; *rec.*, receptor. Proteins with more than 90 residues beyond the end of the first hydrophobic segment are indicated by "°" (see discussion section). One-helix proteins without at least one extramembrane segment longer than 69 residues are indicated by "#" (cf results). Five proteins introduced into the table at a late stage and not included in the statistics and graphs shown elsewhere in this review are indicated by "$".

[b] Plasma membrane proteins whose length or number of putative transmembrane helices is indicated in parentheses: *T-cell receptor β-chain precursor*: The length indicated is that of the unprocessed precursor. *Insulin receptor*: The length indicated is that before cleavage into α and β subunits. *Toll gene product*: Exact position of cleavage of the signal sequence is uncertain; the length indicated is that of the precursor. Notch *gene product*: Exact length uncertain. *CSF-1 receptor* (*c*-fms): Length is approximate, as the exact position of cleavage of the signal sequence is uncertain. *Intestinal sucrase-isomaltase*: The length indicated

is that before cleavage into sucrase and isomaltase. *Stearyl-CoA desaturase*: To our knowledge, no definite transmembrane topological model has been proposed; we have accepted as putative transmembrane segments four regions whose GES was greater than 1.2. *Cytochrome P-450*: Between the 1-helix and 2-helix models proposed by Brown & Black (15), we have favored the 1-helix model because the second helix (residues 36–52) is only moderately hydrophobic, particularly for an anchoring sequence (GES = 0.747; Klein's index P/I = 300). This topology is strongly supported by analysis of the distribution of charges flanking the first helix (53) and by the results of recent proteolysis experiments (139), but some biochemical data are more readily accounted for by the 2-helix model (e.g. 90). *Myelin proteolipid*: Two models have been proposed in which either the first (74) or the third (132) of the four hydrophobic segments forms a hairpin α-helical structure inserted into the outer half of the lipid bilayer. As the topology of the molecule remains uncertain, we have treated each hydrophobic segment as a single transmembrane α-helix. *Na/K ATPase α subunit*: Models with either 8 (126) or 7 (96) hydrophobic transmembrane helices have been proposed. Neurospora *plasma membrane* H^+ *ATPase*: Various topological models have been advanced (1, 50, 124). We have followed Addison (1). Na^+*/glucose co-transporter*: Hediger et al (56) proposed 11 transmembrane helices. Their model, however, neglects several very hydrophobic segments, including one where GES = 2.19 (P/I = 0.4). We have tentatively added those four helices with GES values greater than 1.2. *Voltage-gated* Na^+ *channel* and Ca^{2+} *channel* α_1 *subunit*: Currently accepted models also include four charged transmembrane α-helices, which have not been taken into account in the analysis (92, 135). K^+ *channel* (Shaker *gene product*): The currently accepted model (102) also includes one charged transmembrane α-helix, which has not been taken into account in the analysis.

[c] Mitochondrial inner membrane proteins whose length or number of putative transmembrane helices is indicated in parentheses: *NADH-Q-reductase*: No structural models have been proposed to our knowledge for these subunits. We have accepted as putative transmembrane segments those with a GES ≥ 1.2, except in three cases for which we accepted an additional helix [subunit 5 of *Bos* mitochondrion (GES = 0.96) and subunits 5 and 4L of *Marchantia* chloroplast (GES = 1.13 and GES = 0.91)], which is homologous to more strongly hydrophobic segments in subunits 5 or 4L from other species. *Cytochrome b*: We have followed the revised transmembrane topology proposed by Rao & Argos (111), first discussed in detail by Crofts et al (13, 26, 32c). *ATPase, subunit 6*: A 6-helix structure had been proposed for subunit 6 of *Bos* mitochondrion F_0 and subunit *a* of *E. coli* F_0 (143). More recently, however, subunit IV of chloroplast CF_0 has also been sequenced (see below). Our comparative analysis of these three homologous sequences is more in favor of a 5-helix structure (excluding the fourth and least hydrophobic helix of the 6-helix model). *Succinate-Q reductase, subunits C and D*: These two integral membrane subunits have similar molecular weights in *Bos* mitochondria and *E. coli* but have been sequenced only in *E. coli*. Therefore, hydrophobicity analysis is indicated for *E. coli* subunits. *Nicotinamide nucleotide transhydrogenase (NNT)*: A 14-helix structure had been proposed for *Bos* transhydrogenase, and arguments developed over locating both N and C termini on the matrix side of the membrane (148). Those helices conserved between beef NNT and bacterial NNT [made up of two subunits, α and β (e.g. 22)] are likely oriented in the membrane with the same polarity in the two species. This orientation can be achieved by excluding the fifth putative transmembrane segment of beef NNT 14-helix model, which is located in the link between the two regions homologous to the two bacterial subunits, and has no counterpart in *E. coli*, as well as to the 13th putative helix, which is weakly hydrophobic both in beef (GES = 1.34) and in *E. coli*. *ADP/ATP carrier protein, brown fat uncoupling protein, and phosphate carrier*: These three proteins have a tripartite structure, comprising three similar repeats of approximately 100 residues each. Proposed structures include three hydrophobic transmembrane α-helices (one per repeat) and additional helical or nonhelical amphipathic transmembrane segments (4, 5, 118). Our analysis includes only the hydrophobic helices.

[d] Chloroplast proteins whose length or number of putative transmembrane helices is indicated in parentheses: (Table 4C) *NADH-Q reductase*: See NADH-Q reductase in Footnote b. *ATPase, subunit IV*: a 4-helix structure had been proposed for ATPase subunit IV of *Spinacia* chloroplast (59). We suggest a 5-helix structure including a fifth hydrophobic domain (GES = 2.25), which was previously not considered a transmembrane segment because of its proline content. *Light Harvesting Complex I (LHCI) chlorophyll a/b binding (chlo.ab) proteins, type II and III*: Probable length of the mature proteins has been estimated by analogy with LHCII, whose site of processing has been determined directly (101, 130). (Table 4D) *Phosphate translocator*: The length indicated is that of the precursor. Signal sequence cleavage was suggested from SDS-PAGE analysis to occur around amino acid positions 85–95 (40). The sequence does not contain internal repeats as described for the mitochondrial phosphate carrier, ADP/ATP carrier, and uncoupling carrier (40). (Table 4E) *Hypothetical membrane proteins of chloroplast-encoded ORFs*: Hydrophobic segments were accepted as putative transmembrane α-helices if the probability index P/I of Klein et al (70) was lower than 80 and GES higher than 1.2.

with more than 35–40%-sequence identities to any of those listed have been eliminated. No three-dimensional structure is available for the membrane-embedded region of any of these proteins to a resolution better than about 25 Å, which is too poor to permit identification of individual transmembrane segments. Existing models for transmembrane arrangements are therefore based on a variety of indirect evidence, which ranges from the collection of extensive biochemical data to mere inspection of the hydrophobicity profile generated using any of five different scales (cf 65a).

To estimate the probable number and hydrophobicity of transmembrane segments in each protein in a homogeneous manner, we reanalyzed each sequence using a uniform procedure that searched it for hydrophobic 17-residue segments that could form transmembrane α-helices. The program, derived from that written by Klein et al (70), used two scales, that of Kyte & Doolittle (73) and that of Engelman et al (37) (see the section on procedures). Throughout this chapter, segment hydrophobicities (GES) are expressed in kcal/residue averaged over a 17-residue stretch, using Engelman's scale. In some cases, we also give Klein's P/I index that evaluates the relative probabilities for a segment to be either peripheral or integral.

In most cases, sequence segments proposed in the literature to form hydrophobic transmembrane α-helices overlapped with hydrophobic segments identified by the program. Some segments proposed as transmembrane had only a moderate hydrophobicity, however, and some relatively hydrophobic stretches were found that had been postulated to not span the membrane (Figure 1). As a rule, we accepted as correct the topology proposed in the literature and took as the hydrophobicity of each proposed transmembrane α-helix that of the hydrophobic segment that overlapped it. In 16 cases, however, either the literature offered several models or no model at all, or we felt compelled to not accept the prevailing model. The rationale we have followed in each case is indicated in Footnotes b-d to Table 4. The charged helices thought to make up part of the transmembrane region of voltage-gated channels and presumed to act as voltage sensors have not been included in the analysis, nor have the few nonhelical or strongly hydrophilic transmembrane segments that have been postulated in some models (cf footnotes to Table 4).

Figure 1 shows the overall distribution of GES values for the 589 putative transmembrane α-helices contained in 135 proteins. Table 4 gives GES values for the most hydrophobic and the least hydrophobic of the putative transmembrane segments in each protein, together with the average GES value for the whole transmembrane region of the protein. Some results from this analysis confirm conclusions reached by others (35, 37, 70, 73) using more restricted sets of proteins and a variety of hydro-

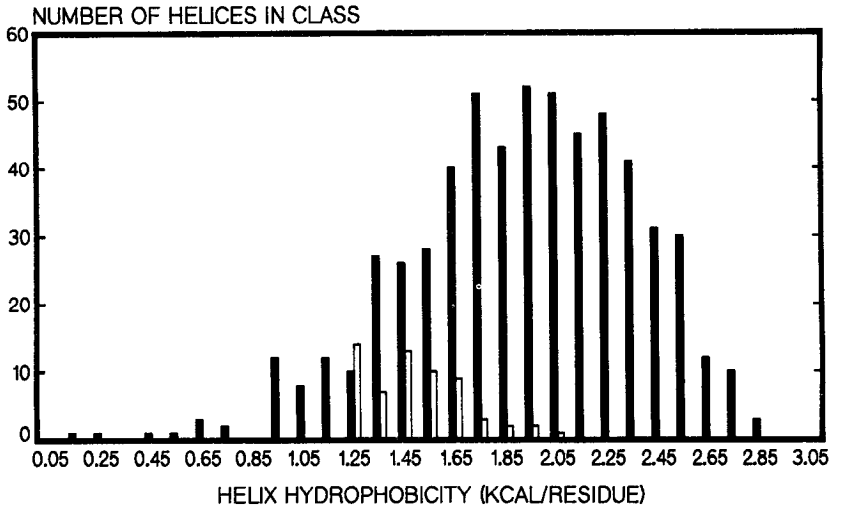


Figure 1 Hydrophobicity distribution of 589 putative transmembrane α-helices (*solid bars*) and 91 presumably extramembrane hydrophobic segments (*open bars*) present in the sequence of 135 integral membrane proteins. Hydrophobic segments corresponding to putative transmembrane α-helices were identified as described in the section on procedures. Their hydrophobicity (free energy cost for transferring them from lipids to water under α-helical configuration) is expressed as GES, the free energy cost per residue averaged over 17-residue stretches using the GES scale [Table 2 (37)]. In the course of the analysis, some rather hydrophobic stretches of residues were found where currently accepted models for the proteins predicted no transmembrane segments. Open bars represent only those whose GES is at least equal to 1.2 kcal/residue.

phobicity scales, namely: 1. In general, from such a crude analysis one can often decide whether a protein is integral or not: 96% of all proteins in Table 4 have at least one segment where GES > 1.5. While we have not analyzed in the same manner an equivalent sample of soluble proteins, we examined most of the presumptive extramembrane domains of the proteins in Table 4 in totality, and they yielded only two dozen segments where GES > 1.5 and one where GES > 2.0 (Figure 1). 2. Once a protein has been classified as integral because it contains at least one very hydrophobic stretch, often one cannot determine with certainty how many other transmembrane stretches are present, as some relatively hydrophobic segments are likely not transmembrane (e.g. in the putative extracellular domain of the *Notch* protein), while some mildly hydrophobic segments are almost certainly transmembrane. Such is the case in G-protein linked receptors. Each most likely has seven transmembrane α-helices, given the overall hydrophobicity pattern throughout the rhodopsin family, but the hydro-

phobicity of some of these segments is quite low (Table 4). In most instances, little or no experimental evidence shows that these less hydrophobic segments actually span the bilayer. Examination of Figure 1 suggests that in the case of totally unknown integral proteins, setting the lowest limit for accepting a segment as transmembrane at 1.2 or 1.3 kcal/residue could result in missing 7–9% of the actual transmembrane segments and mistakenly accepting as transmembrane about the same number of extramembrane ones. This conclusion is similar to that reached by von Heijne (142a) in his analysis of bacterial membrane protein sequences.

Number and Hydrophobicity of Putative Transmembrane α-Helices as a Function of Protein Localization and Function

Figure 2A shows the distribution of eukaryotic integral membrane proteins as a function of length and number of putative transmembrane segments. Plasma membrane and organelle inner membrane proteins differ markedly. Plasma membrane proteins are mainly distributed between anchored proteins, with a single transmembrane helix and large hydrophilic region(s), and polytopic proteins often involved in transmembrane permeation, with many transmembrane helices and rather large extramembrane regions. Extramembrane regions in organelle membrane proteins tend to be smaller. The proteins are more extensively inserted into the bilayer. Some organelle subunits consist of one transmembrane segment and little more. The difference between cellular compartments is particularly striking in Figure 2B, which maps plasma membrane proteins and organelle proteins predicted to span the membrane only once as a function of their length and of the hydrophobicity of the transmembrane helix.

If one considers that about 70 amino acid residues are necessary for a sequence segment to take up a stable three-dimensional conformation, 93% of the one-helix plasma membrane proteins have extramembrane regions large enough to form such domains, against only 33% of the organelle proteins (lack of such regions is indicated by a "#" in Table 4). The latter figure falls to only 9% when unidentified proteins encoded by chloroplast ORFs are included (cf Figure 2C). Most ORFs appear to code for very small proteins barely longer than a single transmembrane α-helix (Figure 2C and Table 2E).

Length discrepancies to some extent reflect the different functions of various membranes. Many plasma membrane proteins interact with molecules located in the extracellular and/or cytoplasmic compartments. Most organelle proteins carry out bioenergetic processes, which involve transmembrane proton or electron transfer, and the binding of lipid-soluble ligands. Differences in function are also reflected in the hydrophobicity

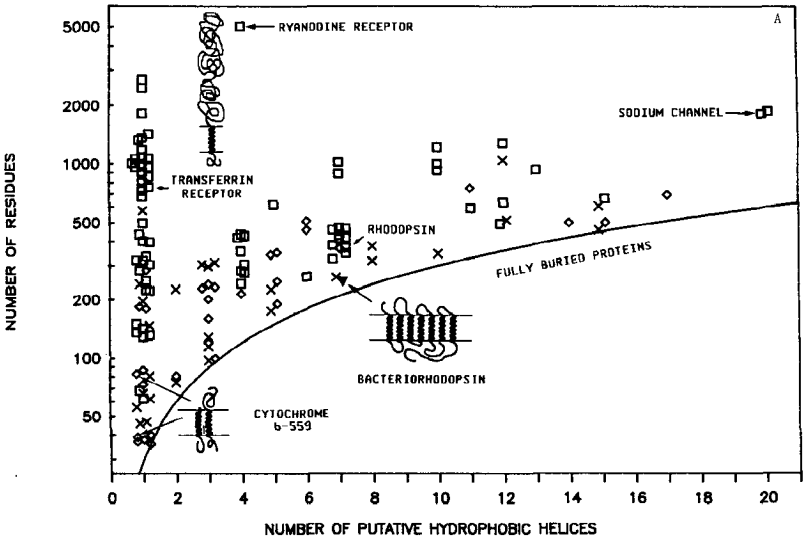
of the transmembrane segment of 1-helix proteins (Figure 2B), which is lower in organelle inner membranes (with a mean GES of 1.9 ± 0.5) than in the plasma membrane (2.3 ± 0.3). Hydrophobicity is particularly high (2.5 ± 0.1) when the α-helix serves as the membrane anchor to a large protein of more than 1000 residues, as are often encountered in plasma membranes.

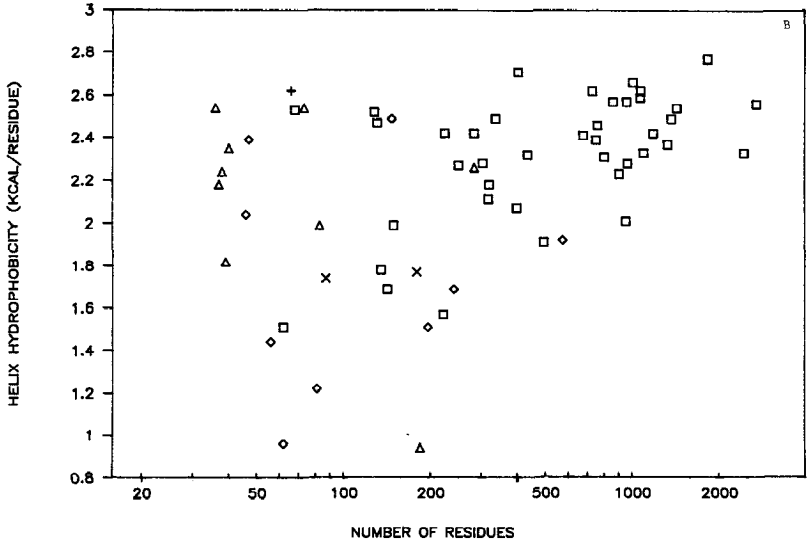
Number and Hydrophobicity of Transmembrane Segments in Organelle Inner-Membrane Proteins Depending on Site of Synthesis

Organelle membrane proteins themselves are not homogeneous. As a rule, proteins with more than three predicted transmembrane segments are encoded by the organelle's DNA (Figure 3). Imported integral membrane proteins (and, in chloroplasts, many plastid encoded ones) have few hydrophobic segments. Furthermore, as shown by Figure 4, imported hydrophobic segments tend to be less hydrophobic than segments synthesized in the organelles (average GES = 1.7 ± 0.4 vs 2.0 ± 0.4). As a result, the total hydrophobicity of the presumptive transmembrane region of proteins imported in the organelles (68 ± 70 kcal) is much lower than for proteins encoded by organelle DNA (188 ± 156 kcal, cf Figure 5). On the other hand, the presence of a highly hydrophobic sequence segment in itself does not prevent importation, e.g. one of the two putative transmembrane α-helices of ATPase subunit 9 (GES = 2.7).

There are two or three exceptions to the tentative rule that proteins with more than three transmembrane segments are not imported. Two are natural (Table 2): Bovine nicotinamide nucleotide transhydrogenase (NNT), an enzyme from mitochondrial inner membrane likely to contain 12 transmembrane segments, is encoded in the nuclear genome (148), as is the phosphate translocator from the inner envelope membrane of chloroplasts [7 putative transmembrane helices, cf (40)]. The third exception is engineered: the chloroplast gene coding for the D1 protein of photosystem II (5 transmembrane segments) from an atrazine-resistant biotype of *Amaranthus hybridus* has been introduced into the nuclear genome of tobacco. Import was inferred from the increased resistance to atrazine of some of the transgenic plants (20).

Differences between mitochondrial genomes from different species suggest a number of other possible exceptions. Genes coding for certain integral membrane proteins are present in some mitochondrial genomes and absent in others (see Table 5). Most strikingly, *Saccharomyces cerevisiae* does not have any mitochondrial genes coding for NADH-Q reductase subunits. *S. cerevisiae* may, however, simply lack this type of NADH dehydrogenase (48). Various forms of reductase indeed exist in different





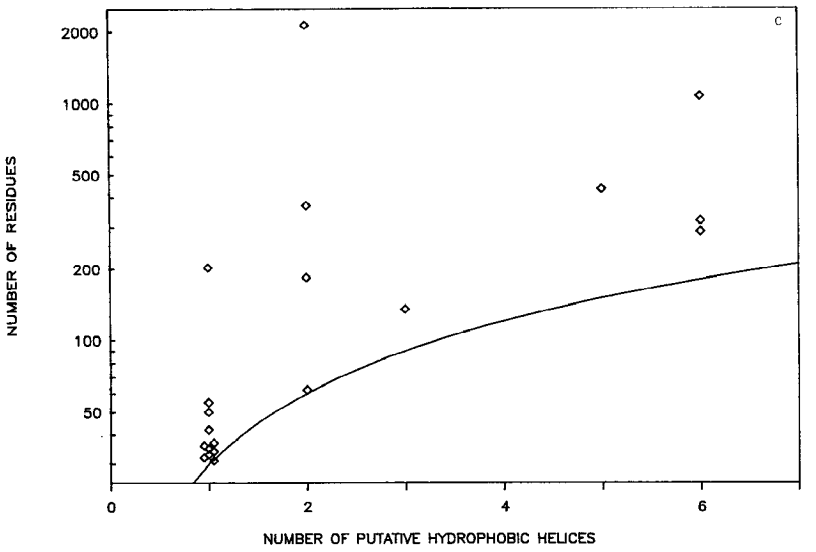


Figure 2 (*A*) Protein length as a function of number of putative hydrophobic transmembrane segments. (*Open squares*) Proteins from membranes that are directly in contact with the cytosol (plasma membrane, endoplasmic and sarcoplasmic reticulum, retina sacculae, exocytotic vesicles). (*Capital X*) Proteins from the inner membrane of mitochondria. (*Open diamonds*) Proteins from the thylakoid membrane [open reading frames (ORFs) are not included]. A well-characterized prokaryotic protein, bacteriorhodopsin, has been added for comparison (*solid triangle*). The curve gives the approximate position of proteins that are essentially fully buried into the bilayer, assuming that approximately 30 residues are needed to span the full thickness of the bilayer (40–45 Å) and form one turn. The folding schemes shown for three proteins illustrate the distribution of mass between putative transmembrane helices and the rest of the protein in different regions of the map; the shape given to the extramembrane regions is arbitrary. The model for the assembly of the two subunits and heme that make up the cytochrome b_{559} heterodimer is taken from Ref. 60. (*B*) Length of proteins spanning the membrane only once (*1-helix proteins*) relative to hydrophobicity of each one's anchoring segment. (*Open squares*) Plasma membrane proteins. (*Plus signs*) Proteins synthesized in mitochondria. (*Open diamonds*) Proteins imported into mitochondria. (*Open triangles*) Proteins synthesized in chloroplasts. (*Capital X*) Proteins imported into chloroplasts. (*C*) Predicted properties of hypothetical proteins encoded by open reading frames (ORFs) in chloroplast genome. The curve has the same meaning as in *A*.

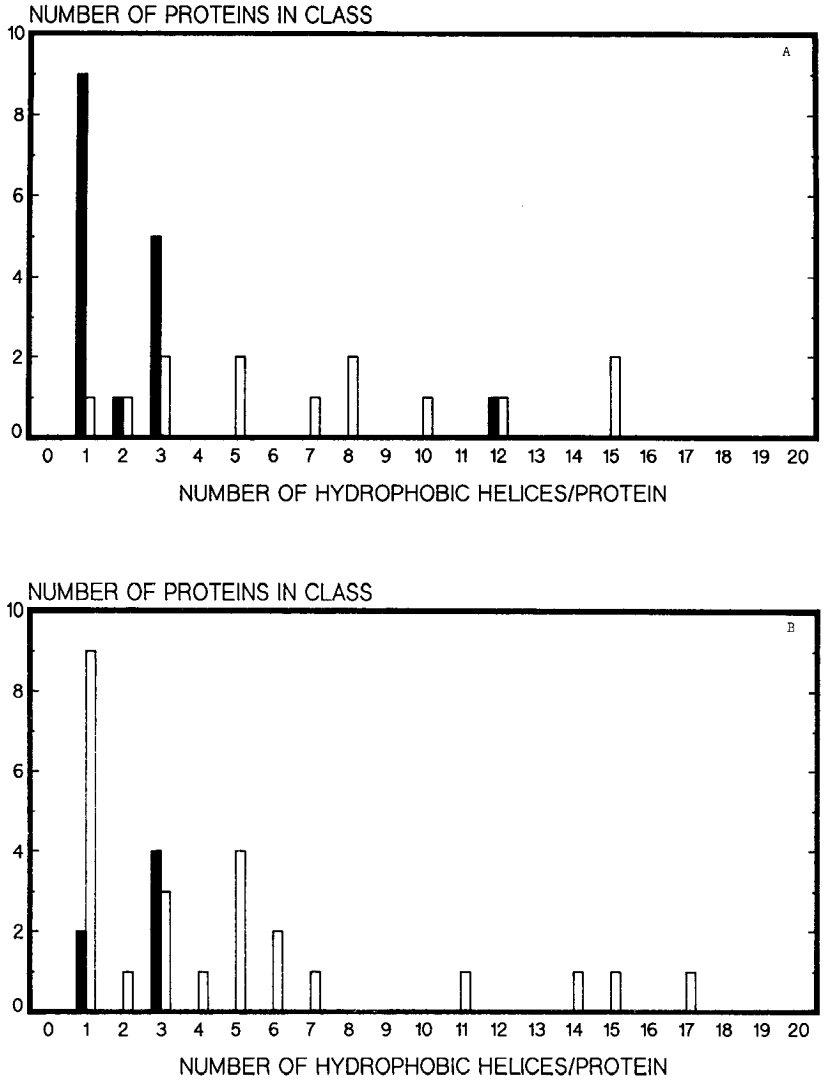


Figure 3 Number of putative hydrophobic transmembrane segments in organelle proteins (excluding ORFs encoding unidentified proteins) depending on whether they are imported from the cytol (*solid bars*) or synthesized in the organelles (*open bars*). (*A*) Mitochondrion. (*B*) Chloroplast.

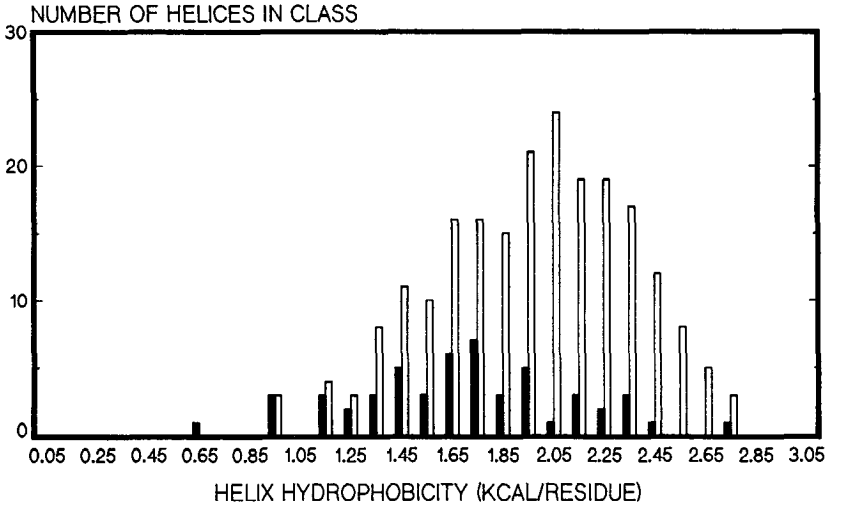


Figure 4 Hydrophobicity distribution of predicted transmembrane segments in organelle proteins depending on site of synthesis. (*Solid bars*) Proteins from mitochondrial inner membrane and thylakoids that are imported from the cytosol. (*Open bars*) Proteins from mitochondrial inner membrane and thylakoids that are synthesized in the organelle (excluding ORFs encoding unidentified proteins).

organisms: in *E. coli*, for instance, the NADH-Q reductase is composed of a single soluble subunit (cf Table 6). In *Leishmania tarentolae* and *Trypanosoma brucei*, some of the genes coding for the proteins of the NADH-Q reductase or the proton ATPase are found in the mitochondrial genome, while others are not [note that extensive editing of mitochondrial mRNA may complicate gene identification (38)]. In these organisms, the missing subunits probably must be imported, which would include subunits 2 and 6 of the reductase (10 and 5 predicted transmembrane segments, respectively, judging from *Bos* sequences) and subunit 6 of the ATPase (5 predicted transmembrane segments). Table 5 lists a number of other subunits whose genes have not been found in largely, but not totally, sequenced mitochondrial genomes from *Chlamydomonas* or *Neurospora*.

DISCUSSION

The Microassembly of Integral Membrane Proteins in Organelles

The analysis presented above bears out the suggestion that integral membrane proteins comprise folding domains that are much smaller than those

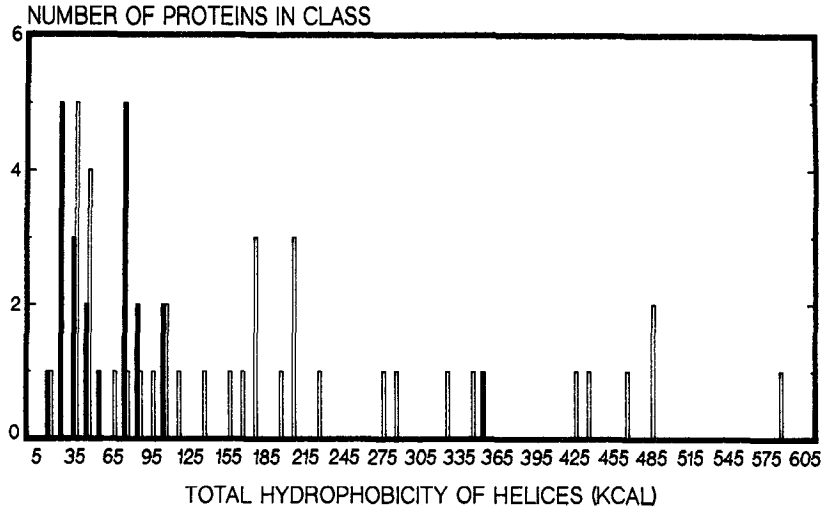


Figure 5 Total hydrophobicity of predicted transmembrane regions in organelle integral proteins depending on site of synthesis. (*Solid bars*) Proteins from mitochondrial inner membrane and thylakoids that are imported from the cytosol. (*Open bars*) Proteins from mitochondrial inner membrane and thylakoids that are synthesized in the organelle (excluding ORFs encoding unidentified proteins). The estimates correspond to the total free energy cost for transferring from lipid to water phase all hydrophobic segments thought to comprise the transmembrane region as a single elongated α-helix. Entropic contributions due to changes in the number of possible geometries for helix association have been neglected.

of soluble proteins. Many of them actually have extramembrane regions that are so small that they would not be expected to fold by themselves into stable structures. Because their hydrophobic environment severely constrains the structure of transmembrane segments, however, individual transmembrane α-helices can provide the specific interactions necessary for assembly with other transmembrane segments belonging to the same and/or other polypeptides. Assembly in turn may constrain the structure of the extramembrane regions. Cytochrome b_{559} provides a good example of the stability of single transmembrane α-helices. This complex is composed of two different small subunits, each probably forming a single transmembrane α-helix (see Table 4C and Figure 2A), and its heme is thought to be liganded by two histidine residues, one in each helix (60). Two copies of this heterodimer are associated with each photosystem II reaction center.

Very low molecular weight integral subunits are mainly noticeable in the complexes from the inner membranes of organelles. How many hydrophobic segments an organelle subunit contains depends on where the

Table 5 Site of synthesis of integral membrane proteins encoded in the mitochondria of at least one species[a]

Species	Genome		NADH-Q reductase							Cyto-chrome *b*	Cytochrome *c* oxidase			ATPase			References
	kbp	%	1	2	3	4	5	6	4L		I	II	III	6	8	9	
Bos tauris	16.5	100	+	+	+	+	+	+	+	+	+	+	+	+	+	−	3
N_h			8	10	3	15	15	5	3	8	12	2	7	5	1	2	
Neurospora crassa	60	80	+	+	?	?	+	+	?	+	+	+	+	+	+	+[b]	14, 63
Saccharomyces cerevisiae	85	92	?	?	?	?	?	?	?	+	+	+	+	+	+	+	32b
Leishmania tarentolae	16.7	100	+	−	−	+	+	−	−	+	+	+	+	−	−	−	128
Trypanasoma brucei	15	100	+	−	−	+	+	−	−	+	+	+	+	−	−	−	38, 128
Chalmydomonas reinhardtii	15.8	80	+	+	?	?	+	+	?	+	+	?	?	?	?	?	45
Higher plants	200–2500	part	+	+	+	?	+	?	?	+	+	+	+	+	+	+	45, 49, 76

[a] The size of the mitochondrial genome in each species (in kilobase pairs) and the proportion of it that has been sequenced are indicated. A plus sign indicates a protein is encoded in the mitochondria; a minus sign indicates it is not encoded in the mitochondria; and a question mark indicates the site of synthesis has not been determined. The number of putative hydrophobic transmembrane segments (N_h) is indicated for the proteins from *Bos*.

[b] *Neurospora* also contains a copy of the ATPase subunit 9 gene in the nucleus.

Table 6 Comparison of integral membrane proteins from eukaryotic organelles and from prokaryotes[a]

6A Mitochondrion vs prokaryote

Complex	Eukaryote & prokaryote		Eukaryote specific		Prokaryote specific	Reference
	Mito.	Nucl.	Mito.	Nucl.		
NADH-Q reductase						151
Bos tauris (vs *E. coli*)						
Subunit 1			8		Only one subunit (47 kDa, 0 helix)	
Subunit 2			10			
Subunit 3			3			
Subunit 4			15			
Subunit 5			15			
Subunit 6			5			
Subunit 4L			3			
Succinate-Q reductase						55, 146
Bos tauris (vs *E. coli*)						
Subunit C		(3)				
Subunit D		(3)				
QH2-Cytochrome *c* Reductase						42
Bos tauris (vs *Rhodobacter sphaeroides*)						
Cytochrome *b*	8					
Cytochrome c_1		1				
FeS protein		1				
Subunit VII				1		
Subunit X				1		
Subunit XI				1		
Cytochrome *c* oxidase						108
Bos tauris (vs *Paracoccus denitrificans*)						
Subunit I	12					
Subunit II	2					
Subunit III	7					
Subunit IV				1		
Subunit VIIIa				1		
Subunit VIIIb				1		
ATPase						81
Bos tauris (vs *E. coli*)						
Subunit 6	5				Subunit b (1 helix)	
Subunit 8			1			
Subunit 9		2				
Nicotinamide nucleotide transhydrogenase						22
Bos tauris (vs *E. coli* subunits α and β)		12 (4+8)				

6B Chloroplast vs prokaryote

Complex	Eukaryote & prokaryote		Eukaryote specific		Prokaryote specific	Reference
	Chlo.	Nucl.	Chlo.	Nucl.		
Photosystem II						
Higher plants (vs *Rhodobacter sphaeroides*)						2
Marchantia psbA	5				Subunit H	
Marchantia psbB			6		(1 helix)	
Marchantia psbC			6			
Marchantia psbD	5					
Nicotiana psbE			1			
Nicotiana psbF			1			
Nicotiana psbH			1			
Marchantia psbI			1			
Marchantia psbJ			1			
Nicotiana psbK			1			
Marchantia psbL			1			
Cytochrome b_6/f						
Marchantia polymorpha (vs *Rhodobacter sphaeroides*)						42
Cytochrome b_6+ Subunit IV (vs *R. sphaeroides* Cytochrome *b*)	4+ 3 (8)					
Cytochrome *f*	1					
FeS protein		1				
Photosystem I						
Algae and higher plants (vs *Chlorobium limicola*)						91
Marchantia P700	11					
Chlamydomonas P37				1		
ATPase						
Marchantia polymorpha (vs *E. coli*)						9, 91
Subunit I	1					
Subunit III	2					
Subunit IV	5					
Antenna					αB870	
Higher plants (vs *Rhodobacter sphaeroides*)					(1 helix) βB870	154
Pisum LHCII CabI				3	(1 helix)	
Lycopersicon LHCI CabI				3	βB800–850	
Petunia LHCI CabII				3	(1 helix)	
Lycopersicon LHCI CabIII				3	βB800–850 (1 helix)	

continued

Table 6 (*continued*)

Complex	Eukaryote & prokaryote		Eukaryote specific		Prokaryote specific	Reference
	Chlo.	Nucl.	Chlo.	Nucl.		
NADH-Q reductase						
Marchantia (vs *E. coli*)						151
Subunit 1			7		Only one	
Subunit 2			15		subunit	
Subunit 3			3		(47 kDa,	
Subunit 4			14		0 helix)	
Subunit 5			17			
Subunit 6			5			
Subunit 4L			3			

[a] Subunits are divided into those common to eukaryotes and prokaryotes, those specific to eukaryotes and those specific to prokaryotes (on the basis of presently available sequences). They are further distributed according to their site of synthesis. The number of helices predicted for each subunit is indicated. *Succinate-Q reductase*: The two integral membrane subunits of this protein have similar molecular weights in *Bos tauris* mitochondria and *E. coli* but have been sequenced only in *E. coli*. The number of helices predicted is indicated for *E. coli* subunits. *Nicotinamide nucleotide transhydrogenase*: This protein is composed of one polypeptide in *Bos tauris* mitochondria and two in *E. coli* (cf Tables 1 and 4B). *Cytochrome b_6/f*: Cytochrome *b* in the QH2-cytochrome *c* reductase complex of *Rhodobacter sphaeroides* corresponds to two subunits [cytochrome b_6 and subunit IV (braced in table)] in cytochrome b_6/f complex (cf Tables 1 and 4C). *ATPase subunit* I: In contrast to ATPase subunit 8 of mitochondria, subunit I of chloroplast CF_0 shows some similarity in primary structure to subunit *b* of *E. coli* F_0. *ATPase subunit IV*: Regarding the number of putative transmembrane α-helices, see Footnote d for Table 4.

subunit is synthesized. In mitochondrion inner membrane, there is a clear-cut discrepancy: with few exceptions, subunits are imported if they contain three or fewer putative transmembrane α-helices and are synthesized in situ if they contain more than three. In chloroplast thylakoids, imported proteins also have few hydrophobic segments, but proteins synthesized in situ can have either many or few. These distributions do not indirectly result from a tendency for large genes to remain in the organelles, since most soluble or extrinsic proteins, whether large or small, are synthesized in the cytoplasm (3, 93, 125). The exclusion is not absolute; there exist at least two natural and one engineered exceptions. These are discussed below. Nevertheless, the tendency of imported proteins to have few transmembrane segments is very strong.

The causes of this distribution are uncertain, and may involve several factors. The possibility of mistargeting was proposed by von Heijne (141a, 142), who argued that if a hydrophobic segment appeared in the cytosol before the synthesis of a nuclear-encoded protein is completed, it would

act as a signal sequence and target the protein to the endoplasmic reticulum. This could occur if there are more than 70–90 residues after the end of the first hydrophobic segment. More recent data, however, suggest that mistargeting cannot be a decisive factor because, at least under this simple form, this hypothesis predicts the misdirection of about half of the imported proteins (marked with a degree sign in Table 4).

Other possible explanations might involve the mechanism of import into the mitochondrion or chloroplast. Considerable evidence indicates that import is, primarily or totally, posttranslational (for reviews, see 6, 52, 106, 141). Import involves unfolding of the protein to be translocated. It is prevented by stabilization of the mature, folded conformation (18, 31, 33, 121). Cytosolic proteins are involved in preventing folding or aggregation of the nascent chains and/or in unfolding the chains into an import-competent form (e.g. 32, 34, 51, 52, 98, 100, 117). Similar proteins play a role in protein folding and assembly in the mitochondrial matrix (19, 95, 112).

The presence of a large number of hydrophobic residues in a polypeptide can perturb import in several ways. To prevent aggregation and precipitation or nonspecific association with membranes or with other proteins, large hydrophobic patches must not be exposed to the cytosol. This can be achieved either by appropriate folding of the protein or by association with itself or with other proteins. Particular difficulties are expected for integral membrane proteins because, in contrast to soluble proteins, they expose a considerable hydrophobic area to their surface in their native state. The achievement of a soluble conformation or complex should become increasingly more difficult as the number of hydrophobic segments to be masked increases. Problems also might arise at the unfolding step, since hydrophobic residue burial is a major source of stabilization free energy for folded structures and for oligomers. The more hydrophobic residues have been buried, the more difficult it is to unfold or dissociate the resulting structure. Similar difficulties may also be encountered on the matrix or stroma side of the organelles.

The recent description of a soluble form of the integral protein *lac* permease (116) or, in contrast the hydrophobic properties of bacterial porins can serve as reminders that the behavior of membrane proteins does not always match our expectations. To surmise that large hydrophobic peptides may stand greater chances of misfolding and precipitating before being targeted to the organelle, or may become untractably difficult to unfold seems reasonable, however. Another conceivable difficulty is the probable tendency of hydrophobic segments located away from the region being translocated to insert nonspecifically into nearby membranes. It is probably significant that imported segments tend to be less hydrophobic

than those in subunits synthesized in the organelles (even though the two distributions overlap). This tendency further decreases the total hydrophobicity of the putative transmembrane region in imported proteins. The involvement of hydrophilic regions, e.g. presequences, in stabilizing the precursor form of imported integral proteins has been discussed by Hartl et al (52).

Borst (12) seems to have first proposed the idea that the biosynthesis of some proteins within organelles rather than in the cytosol could be linked to their hydrophobicity. This view has remained in relative disfavor, presumably because the interest in the process of translocation itself has focused attention on local properties of the sequence. The example of ATPase subunit 9 shows that local hydrophobicity in imported proteins can be very high (Table 4B). The critical importance of the unfolding step, however, may explain why an accumulation of sequence segments that individually would be importable might have an inhibitory effect.

The tentative rule that proteins with more than three hydrophobic transmembrane segments are not imported presently has two natural exceptions. One is nicotinamide nucleotide transhydrogenase, an enzyme from the inner mitochondrial membrane. The other is the phosphate translocator from the inner envelope membrane of chloroplasts. NNT contains probably 12 and the translocator 7 hydrophobic segments (cf Table 4) whose hydrophobicity is typical of that of imported proteins. It might be interesting to determine whether the import of these proteins is posttranslational or coupled to their synthesis.

Whatever the reason(s), in most cases organelles do not import proteins with large transmembrane regions. Eukaryotic cells generally have supplemented complexes inherited from the original symbiotic prokaryotes with additional subunits. In the mitochondrial respiratory chain, all of the new material encoded in the nucleus is made up of 1-helix subunits (Table 6). In addition, the genes for some of the smaller, 1–3-helix integral subunits of prokaryotic origin have been displaced to the nucleus. In chloroplasts, the imported material is made up of 1–3-helix proteins; most of the new subunits are locally encoded and can have either many or few hydrophobic segments. Many of the hypothetical proteins encoded by ORFs are predicted to be small, 1-helix proteins. The existence of many 1-helix subunits encoded in plastid DNA indicates that restriction on import is not the only circumstance in which one encounters such subunits. The building up of organelle complexes seems to take full advantage of the domainlike behavior of transmembrane α-helices: they are put together in a piecemeal manner by a process of microassembly that uses numerous small subunits in addition to a few large ones.

Other Membranes

Very small 1-helix integral subunits are rare among plasma membrane proteins (2 in our sample of 80). Part of the reason for this near absence could be methodologic (such proteins migrate with the dye front in most commonly used SDS-PAGE systems), and part is certainly linked to the different functions of this membrane (see section on results). Differences in biosynthesis may also play a role. It is not clear yet to what extent insertion of proteins into the endoplasmic reticulum is co- or post-translational (for recent discussion, see 43). Similarities between insertion into the ER and import into organelles are certainly greater than previously recognized (for reviews, see 106, 141). For instance, evidence of a role for stress proteins in yeast has recently been obtained (21, 32, 153). However, translation and insertion seem more closely coupled in the ER than for organelle proteins. The red blood cell glucose transporter, a protein thought to contain 12 transmembrane segments, can be imported post-translationally into dog pancreas microsomes, albeit with a low efficiency. Cotranslational insertion or engineered shortening of the polypeptide by 4 transmembrane segments increases efficiency (85).

There are conceivable advantages to using several small subunits instead of a single large one from, for example, evolutionary or regulatory points of view. As mentioned above, numerous small 1-helix proteins are synthesized in situ in chloroplasts. Restriction to import cannot be the reason for their abundance. The scarcity of very small plasma membrane proteins is likely due in part to the absence of the restriction on helix number that seems to be associated with posttranslational import and insertion. It may also be that, in the plasma membrane, any other potential advantage of microassembly is offset by the greater instability of complexes as compared with single-chain proteins or by the increased complexity of targeting and assembly.

We have examined under identical conditions the protein composition of another membrane toward which export of hydrophobic proteins is known to present difficulties, namely the outer membrane of gram-negative bacteria (for reviews, see 7, 106, 107, 110). Export or membrane integration of proteins in bacteria presents similarities with import into organelles in that it can be posttranslational (145); it is prevented by stabilization of the mature, folded form (97, 109) and it involves ATP-dependent antifolding proteins (23, 27, 28, 74a).

We have analyzed the sequences of 11 outer membrane proteins, none of which is thought to form transmembrane hydrophobic α-helices (Table 7). In agreement with the literature, most contained no significantly hydro-

Table 7 Hydrophobicity analysis of integral proteins from the outer membrane of *E. coli*[a]

Protein	Nr	GES	Reference
Murein-lipoprotein	58	−1.28	NBRF
Phospholipase A	260	0.60	NBRF
OmpF (porin)	340	0.34	NBRF
OmpC (porin)	346	0.21	83
PhoE	330	0.42	NBRF
OmpA	325	0.85	NBRF
TolC	467	1.27	NBRF
FhuA (TonA)	714	1.27	NBRF
BtuB	594	0.95	58
LamB (maltoporin)	421	0.36	NBRF
Lc	342	0.39	NBRF

[a] The number of residues (Nr) and the hydrophobicity of the most hydrophobic 17-residue segment (GES; kcal/residue) are given for the mature protein.

phobic segments at all, and two contained a mildly hydrophobic segment with a GES barely higher than 1.2. That the bacterial cell has difficulties exporting hydrophobic proteins is directly substantiated by experiments in which stretches of hydrophobic residues were introduced genetically into the sequence of either a viral coat protein, the natural anchoring segment of which had been deleted (30), or an outer membrane protein (80). In both cases, export was blocked as the length of the hydrophobic insert increased. We have estimated the local hydrophobicity of the 17-residue segments that included these inserts, using the same procedure as for natural proteins. Within some variability, segments with GES values less than 1.6 allowed export and segments with GES values higher than approximately 2.2 blocked it. Partial exportation was observed between these two limits. On this basis, only half a dozen of the 140 proteins listed in Table 4 could conceivably be exported efficiently by *E. coli*.

Porins are the best known outer membrane proteins. They have rather polar sequences and are known to be essentially comprised of β sheets (for recent review, see 8). A strong restriction on the export of hydrophobic segments may explain why the structural solution adopted by porins differs from the microassembly observed in organelle complexes. On the other hand, factors such as the peculiar structure, function and environment of bacterial outer membranes should not be forgotten. The sequence of the porin from yeast mitochondrial outer membrane—a protein apparently

unrelated to bacterial porins (82a)—is also fairly hydrophilic. The GES value of its most hydrophobic 17-residue segment is only 0.72.

Displacing the Synthesis of Integral Membrane Proteins from Organelles to the Cytoplasm

The observations summarized here may shed light on the conditions under which a protein synthesis can be displaced from organelle to cytoplasm, either during the course of evolution (see 46) or as the result of genetic engineering (cf 36).

As already mentioned, proteins like NNT or the chloroplast envelope phosphate translocator appeared atypical in our analysis. Perhaps their biosynthesis presents peculiarities—for instance a closer coupling between translation and import. One can also wonder whether differences exist between homologous proteins depending on their site of synthesis. For example, some complex I subunits synthesized in situ in mammalian mitochondria are presumably imported from the cytoplasm in the parasitic protozoa *Leishmania* and *Trypanosoma* (Table 5). Does the average hydrophobicity of the transmembrane segments in imported segments diminish? Are proteins with many transmembrane segments split into several smaller ones?

Only a few cases of displaced subunits can presently be analyzed from this point of view. Within eukaryotes, comparison of a subunit imported from the cytosol with an equivalent one synthesized in situ is possible for ATPase subunit 9 and for cytochrome c_1. The average hydrophobicity of the two transmembrane segments of ATPase subunit 9 is similar whether the protein is encoded in the nucleus (as in mammals and *Neurospora*), in the mitochondrion (as in yeast and maize), or in the chloroplast. In contrast, the hydrophobicity of the putative transmembrane helix of mitochondrial cytochrome c_1, which is encoded in the nucleus, is much lower than the hydrophobicity of the equivalent segment in cytochrome *f*, which is encoded in the chloroplast (respective GES values 1.53 and 2.26).

Comparison of eukaryotic and prokaryotic complexes shows that the following integral subunits have been displaced to the eukaryote nucleus (Table 6): cytochrome c_1 and FeS subunits of the QH2-cytochrome *c* reductase complex, ATPase subunit 9, and NNT. Again, the hydrophobicity of the anchoring sequence of cytochrome c_1 is found to be much higher when it is not imported (GES = 2.24 in *E. coli* vs 1.53 in eukaryotes). In the other cases, the hydrophobicity remains about the same regardless of import. It is probably premature to draw conclusions from such a limited comparison, particularly as it does not include 3-helix proteins, which presumably would be most sensitive to selective pressure on their hydrophobicity.

We have as yet no examples of a protein with many transmembrane segments that would be split into several smaller ones when its structural gene is displaced to the nucleus. We are aware of only two natural cases of split integral proteins (Table 1), if one leaves aside voltage-gated channels from the plasma membrane, in which a homooligomer in one case (K^+ channel) appears to correspond to a single polypeptide with internal repeats in others (Na^+ and Ca^{2+} channels; cf Table 4A). In the case of cytochrome *b*, neither the whole polypeptide nor the fragments need be imported. In case of NNT, the fragments are not imported while the full-length protein is.

The existence of restrictions to integral protein import can be experimentally tested by displacing the locus of synthesis of organelle-encoded proteins to the cytoplasm. Nagley et al (87) found nucleus-encoded subunit 8 of the F_0 ATPase (fused to a mitochondrial targeting peptide) to rescue yeast mutants lacking functional mitochondrial subunit 8. This observation does not test the ideas developed here, as subunit 8 is a very short protein (48 amino acid residues in yeast) purported to comprise a single transmembrane segment (138). Its structural gene is absent from *Xenopus* mitochondrial genome (115), which suggests that in this organism it is naturally encoded in the nucleus.

Such is not the case for D1, the photosystem II quinone-binding protein that carries the site of action of the herbicide atrazine. D1 is encoded by the *psbA* gene, which is present in every chloroplast genome sequenced thus far (93, 125). This protein most likely features five transmembrane segments (82, 137). Cheung et al (20) have reported that introduction of the *psbA* gene from an atrazine-resistant biotype into the tobacco nuclear genome conferred an increased tolerance to atrazine to some of the transformed plants. This observation suggests that the existence of an absolute barrier to the import of D1 is not the reason for retention of the *psbA* gene in the chloroplast. The efficiency of the import was not established directly and was difficult to assess from functional data because the engineered protein had to compete with the natural one whose synthesis was not blocked. Such a competition could explain the limited resistance to atrazine of the engineered strains.

Further experiments are needed to establish to what extent efficient import can be achieved for multispanning proteins. Our data suggest that low yields of import may be encountered. One attractive possibility for molecular genetic experiments involves splitting genes coding for polytopic proteins into two or more smaller parts, each preceded by a segment coding for an organelle targeting peptide. It is not unreasonable to expect that the resulting protein fragments could assemble in the organelle inner membrane into functional complexes.

CONCLUSION

The present analysis supports the idea that transmembrane α-helices represent autonomous folding domains in integral membrane proteins. It further suggests the existence of biosynthetic problems associated with the posttranslational import or export of proteins containing long stretches of hydrophobic residues. In organelles, restrictions to import are not absolute, and these problems are circumvented by importing numerous small subunits containing few hydrophobic segments. These are subsequently microassembled into complexes thanks to the domainlike behavior of transmembrane α-helices. In the endoplasmic reticulum, microassembly is presumably not required and is in fact seldom observed.

ACKNOWLEDGMENTS

We are grateful to D. M. Engelman, J.-P. Henry, P. Joliot, M. Le Maire, W. Neupert, F. Pattus, M. Uzan and G. von Heijne for comments on the manuscript and discussions, to the late P. Klein for a copy of the source code of the program described in Ref. 70 and to D. Béal for his help with the graphic system used to generate the figures. Sequence retrieval and homology analyses were performed using computer facilities at the CITI2 with the help of the Ministère de la Recherche et de la Technologie. This work was supported by a grant from the Ministère de la Recherche et de la Technologie to J.-L. Popot.

Literature Cited

1. Addison, R. 1986. *J. Biol. Chem.* 261: 14896–14901
2. Allen, J. P., Feher, G., Yeates, T. O., Komiya, H., Rees, D. C. 1987. *Proc. Natl. Acad. Sci. USA* 84: 6162–66
3. Anderson, S., de Bruijn, M. H. L., Coulson, A. R., Eperon, I. C., Sanger, F., Young, I. G. 1982. *J. Mol. Biol.* 156: 683–717
4. Aquila, H., Link, T. A., Klingenberg, M. 1985. *EMBO J.* 4: 2369–76
5. Aquila, H., Link, T. A., Klingenberg, M. 1987. *FEBS Lett.* 212: 1–9
6. Attardi, G., Schatz, G. 1989. *Annu. Rev. Cell Biol.* 4: 289–333
7. Baker, K., Mackman, N., Holland, I. B. 1987. *Prog. Biophys. Molec. Biol.* 49: 89–115
8. Benz, R., Bauer, K. 1988. *Eur. J. Biochem.* 176: 1–19
9. Bird, C. R., Koller, B., Auffret, A. D., Huttly, A. K., Howe, C. J., et al. 1985. *EMBO J.* 4: 1381–88
10. Blank, U., Ra, C., Miller, L., White, K., Metzger, H., Kinet, J.-P. 1989. *Nature* 337: 187–89
11. Borchart, U., Madleidt, W., Schägger, H., Link, T. A., von Jagow, G. 1986. *FEBS Lett.* 200: 81–86
12. Borst, P. 1977. *TIBS* 2: 31–34
13. Brasseur, R. 1988. *J. Biol. Chem.* 263: 12571–75
14. Breitenberger, C. A., LajBhandary, U. L. 1985. *TIBS* 10: 478–82
15. Brown, C. A., Black, S. D. 1989. *J. Biol. Chem.* 264: 4442–49
16. Bunzow, J. R., Van Tol, H. H. M., Grandy, D. K., Albert, P., Salon, J., et al. 1987. *Nature* 783: 336–786
17. Cashmore, A. R. 1984. *Proc. Natl. Acad. Sci. USA* 81: 2960–64
18. Chen, W.-J., Douglas, M. G. 1987. *J. Biol. Chem.* 262: 15605–9
19. Cheng, M. Y., Hartl, F.-U., Martin,

J., Pollock, R. A., Kalousek, F., et al. 1989. *Nature* 337: 620–25
20. Cheung, A. Y., Bogorad, L., Van Montagu, M., Schell, J. 1988. *Proc. Natl. Acad. Sci. USA* 85: 391–95
21. Chirico, W. J., Waters, M. G., Blobel, G. 1988. *Nature* 332: 805–10
22. Clarke, D. M., Loo, T. W., Gillam, S., Bragg, P. D. 1986. *Eur. J. Biochem.* 158: 647–53
23. Collier, D. N., Bankaitis, V. A., Weiss, J. B., Bassford, P. J. Jr. 1988. *Cell* 53: 273–83
24. Collins, J. H., Leszyk, J. 1987. *Biochemistry* 26: 8665–68
25. Coussens, L., Van Beveren, C., Smith, D., Chen, D., Chen, E., Mitchell, R. L., et al. 1986. *Nature* 320: 277–80
26. Crofts, A., Robinson, H., Andrews, K., Van Doren, S., Berry, E. 1987. In *Cytochrome Systems. Molecular Biology & Bioenergetics*, ed S. Papa, B. Chance, L. Ernster, pp. 617–31. New York/London: Plenum
27. Crooke, E., Guthrie, B., Lecker, S., Lill, R., Wickner, W. 1988. *Cell* 54: 1003–11
28. Crooke, E., Wickner, W. 1987. *Proc. Natl. Acad. Sci. USA* 84: 5216–20
29. Cunningham, B. A., Hemperly, J. J., Murray, B. A., Prediger, E. A., Brackenbury, R., Edelman, G. M. 1987. *Science* 236: 799–806
30. Davis, N. G., Model, P. 1985. *Cell* 41: 607–14
30a. Deisenhofer, J., Michel, H. 1989. *Science* 245: 1463–73
31. della-Cioppa, G., Kishore, G. 1988. *EMBO J.* 7: 1299–1305
32. Deshaies, R. J., Koch, B. D., Werner-Washburne, M., Craig, A. A., Schekman, R. 1988. *Nature* 332: 801–5
32a. de Vitry, C., Popot, J.-L. 1989. *C. R. Acad. Sci. (Paris) Ser. 3* 309: 709–14
32b. de Zamaroczy, M., Bernardi, G. 1986. *Gene* 47: 155–77
32c. di Rago, J.-P., Colson, A. M. 1988. *J. Biol. Chem.* 263: 12564–70
33. Eilers, M., Schatz, G. 1986. *Nature* 322: 229–32
34. Eilers, M., Schatz, G. 1988. *Cell* 52: 481–83
35. Eisenberg, D., Schwarz, E., Komaromy, M., Wall, R. 1984. *J. Mol. Biol.* 179: 125–42
36. Ellis, J. 1985. *Nature* 313: 353–54
37. Engelman, D. M., Steitz, T. A., Goldman, A. 1986. *Annu. Rev. Biophys. Biophys. Chem.* 15: 321–53
38. Feagin, J. E., Abrahan, J. M., Stuart, K. 1988. *Cell* 53: 413–22
39. Finkelstein, A. V., Ptitsyn, O. B. 1987. *Prog. Biophys. Mol. Biol.* 50: 171–90
40. Flügge, U. I., Fischer, K., Gross, A., Sebald, W., Lottspeich, F., Eckerskorn, C. 1989. *EMBO J.* 8: 39–46
41. Franzén, L. G., Frank, G., Zuber, H., Rochaix, J.-D. 1989. *Mol. Gen. Genet.* 219: 137–44
41a. Faruichi, T., Yoshikama, S., Miyawaki, A., Wada, K., Maeda, N., Mikoshiba, K. 1989. *Nature* 342: 32–38
42. Gabellini, N., Sebald, W. 1986. *Eur. J. Biochem.* 154: 569–79
43. Garcia, P. D., Walter, P. 1988. *J. Cell Biol.* 106: 1043–48
44. Ghelis, C., Yon, J. 1982. *Protein Folding* New York: Academic. 562 pp.
45. Gray, M. W., Boer, P. H. 1988. *Phil. Trans. R. Soc. Lond.* B319: 135–47
46. Gray, M. W., Doolittle, W. F. 1982. *Microbiol. Rev.* 46: 1–42
47. Grenningloh, G., Rienitz, A., Schmitt, B., Methfessel, C., Zensen, M., et al. 1987. *Nature* 328: 215–20
48. Grivell, L. A. 1989. *Eur. J. Biochem.* 182: 477–93
49. Gualberto, J. M., Wintz, H., Weil, J.-H., Grienenberger, J.-M. 1989. *Mol. Gen. Genet.* 215: 118–27
50. Hager, K. M., Mandala, S. M., Davenport, J. W., Speicher, D. W., Benz, E. J. Jr., Slayman, C. W. 1986. *Proc. Natl. Acad. Sci. USA* 83: 7693–97
51. Harmey, M. A., Neupert, W. 1985. *The Enzymes of Biological Membranes*, Vol. 4, ed. A. Martonosi, pp. 431–64. New York: Plenum
52. Hartl, F.-U., Pfanner, N., Nicholson, D. W., Neupert, W. 1989. *Biochim. Biophys. Acta* 988: 1–45
53. Hartmann, E., Rapoport, T. A., Lodish, H. F. 1989. *Proc. Natl. Acad. Sci. USA* 86: 5786–90
54. Hashimoto, C., Hudson, K. L., Anderson, K. V. 1988. *Cell* 52: 259–79
55. Hatefi, Y. 1985. *Annu. Rev. Biochem.* 54: 1015–69
56. Hediger, M. A., Coady, M. J., Ikeda, T. S., Wright, E. M. 1987. *Nature* 330: 379–81
57. Heinemeyer, W., Alt, J., Herrmann, R. G. 1984. *Curr. Genet.* 8: 543–49
58. Heller, K., Kadner, R. J. 1985. *J. Bacteriol.* 161: 904–8
59. Hennig, J., Herrmann, R. G. 1986. *Mol. Gen. Genet.* 203: 117–28
60. Herrmann, R. G., Alt, J., Schiller, B., Widger, W. R., Cramer, W. A. 1984. *FEBS Lett.* 176: 239–44
61. Hoffman, N. E., Pichersky, E., Malik, V. S., Castrana, C., Ko, K., et al. 1988. *Proc. Natl. Acad. Sci. USA* 84: 8844–48
62. Huang, K.-S., Bayley, H., Liao, M.-J., London, E., Khorana, H. G. 1981. *J. Biol. Chem.* 256: 3802–9

63. Ise, W., Haiker, H., Weiss, H. 1985. *EMBO J.* 4: 2075–80
64. Jaenicke, R. 1987. *Prog. Biophys. Molec. Biol.* 49: 117–237
65. Janin, J., Wodak, S. J. 1983. *Prog. Biophys. Molec. Biol.* 42: 21–78
65a. Jennings, M. L. 1989. *Annu. Rev. Biochem.* 58: 999–1027
66. Julius, D., MacDermott, A. B., Axel, R., Jessell, T. M. 1988. *Science* 241: 558–64
67. Jund, R., Weber, E., Chevallier, M.-R. 1988. *Eur. J. Biochem.* 171: 417–24
68. Kahn, T. W., Adair, B. D., Sigrist, H., Engelman, D. M. 1989. *Biophys. J.* 55: 126a
69. Kahn, T. W., Engelman, D. M. In preparation
70. Klein, P., Kanehisa, M., DeLisi, C. 1985. *Biochim. Biophys. Acta* 815: 468–76
71. Kobilka, B. K., Kobilka, T. S., Daniel, K., Regan, J. W., Caron, M. G., Lefkowitz, R. J. 1988. *Science* 240: 1310–16
72. Krupinski, J., Coussen, F., Bakalyar, H. A., Tang, W. J., Feinstein, P. G., et al. 1989. *Science* 244: 1558–64
73. Kyte, J., Doolittle, R. F. 1982. *J. Mol. Biol.* 157: 105–32
73a. Lalanne, J.-L., Transy, C., Guérin, S., Dorche, S., Meulien, P., et al. 1985. *Cell* 41: 469–78
74. Laursen, R. A., Samiullah, M., Lees, M. B. 1984. *Proc. Natl. Acad. Sci. USA* 81: 2912–16
74a. Lecker, S., Lill, R., Ziegelhoffer, T., Georgopoulos, C., Bassford, P. J. Jr., et al. 1989. *EMBO J.* 8: 2703–9
75. Leube, R. E., Kaiser, P., Seiter, A., Zimbelmann, R., Franke, W. W., et al. 1987. *EMBO J.* 6: 3261–68
76. Levings, C. S. III, Brown, G. G. 1989. *Cell* 56: 171–79
77. Liao, M.-J., Huang, K.-S., Khorana, H. G. 1984. *J. Biol. Chem.* 259: 4200–4
78. Liao, M.-J., London, E., Khorana, H. G. 1983. *J. Biol. Chem.* 258: 9949–55
79. Loosfelt, H., Misrahi, M., Atger, M., Salesse, R., Vu Hai-Luu Thi, M. T., et al. 1989. *Science* 245: 525–28
80. MacIntyre, S., Freudl, R., Eschbach, M.-L., Henning, U. 1988. *J. Biol. Chem.* 263: 19053–59
81. Maloney, P. C. 1987. Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology*, Vol. 17, ed. F. C. Neidhart, pp. 222–43. Washington DC: American Society for Microbiology
82. Michel, H., Deisenhofer, J. 1988. *Biochemistry* 27: 1–7
82a. Mihara, K., Sato, R. 1985. *EMBO J.* 4: 769–74
83. Mizuno, T., Chou, M.-Y., Inouye, M. 1983. *J. Biol. Chem.* 258: 6932–40
84. Morgan, D. O., Edman, J. C., Standring, D. N., Fried, V. A., Smith, M. C., et al. 1987. *Nature* 329: 301–7
85. Mueckler, M., Lodish, H. F. 1986. *Cell* 44: 629 37
86. Murata, N., Miyao, M., Hayashida, N., Hidaka, T., Sugiura, M. 1988. *FEBS Lett.* 235: 283–88
87. Nagley, P., Farrell, L. B., Gearing, D. P., Nero, D., Meltzer, S., Devenish, R. J. 1988. *Proc. Natl. Acad. Sci. USA* 85: 2091–95
88. Nakano, Y., Guerrero, I., Hidalgo, A., Whittle, J. R. S., Ingham, P. W. 1989. *Nature* 341: 508–13
89. Needleman, S. B., Wunsch, C. D. 1970. *J. Mol. Biol.* 48: 443–53
90. Nelson, D. R., Strobel, H. W. 1988. *J. Biol. Chem.* 263: 6038–50
91. Nitschke, W., Feiler, U., Lochau, W., Hauska, G. 1987. *FEBS Lett.* 218: 283–86
92. Noda, M., Shimizu, S., Tanabe, T., Takai, T., Kayano, T., et al. 1984. *Nature* 312: 121–27
93. Ohyama, K., Fukuzawa, H., Kohchi, I., Shirai, H., Sano, T., et al. 1986. *Nature* 322: 572–74
94. Olsen, J., Cowell, G. M., Königshöfer, E., Danielsen, E. M., Møller, J., et al. 1988. *FEBS Lett.* 238: 307–14
95. Ostermann, J., Horwich, A. L., Neupert, W., Hartl, F.-U. 1989. *Nature* 341: 125–30
96. Ovchinnikov, Yu. A., Luneva, N. M., Arystarkhova, E. A., Gevondyan, N. M., Arzamazova, N. M., et al. 1988. *FEBS Lett.* 227: 230–34
97. Park, S., Liu, G., Topping, T. B., Cover, W. H., Randall, L. L. 1988. *Science* 239: 1033–35
98. Pelham, H. R. B. 1986. *Cell* 46: 959–61
99. Peralta, E. G., Winslow, J. W., Peterson, G. L., Smith, D. H., Askhenazi, A., et al. 1987. *Science* 236: 600–5
100. Pfanner, N., Hartl, F.-U., Neupert, W. 1988. *Eur. J. Biochem.* 175: 205–12
101. Pichersky, Y. E., Brock, T. G., Nguyen, D., Hoffman, N. E., Piechulla, B., et al. 1989. *Plant. Mol. Biol.* 12: 257–70
102. Pongs, O., Kecskemethy, N., Müller, R., Krah-Jentgens, I., Baumann, A., et al. 1988. *EMBO J.* 7: 1087–96
103. Popot, J.-L., Engelman, D. M. 1990. *Biochemistry*. In press
104. Popot, J.-L., Gerchman, S.-E., Engelman, D. M. 1987. *J. Mol. Biol.* 198: 655–76

105. Popot, J.-L., Trewhella, J., Engelman, D. M. 1986. *EMBO J.* 5: 3039–44
106. Pugsley, A. P. 1989. *Protein Targeting.* San Diego: Academic. 279 pp.
107. Pugsley, A. P., Schwartz, M. 1985. *FEMS Microbiol. Rev.* 32: 3–38
108. Raitio, M., Jalli, T., Saraste, M. 1987. *EMBO J.* 6: 2825–33
109. Randall, L. L., Hardy, S. J. S. 1986. *Cell* 46: 921–28
110. Randall, L. L., Hardy, S. J. S., Thom, J. R. 1987. *Annu. Rev. Microbiol.* 41: 507–41
111. Rao, J. K., Argos, P. 1986. *Biochim. Biophys. Acta* 896: 197–205
112. Reading, D. S., Hallberg, R. L., Myers, A. M. 1989. *Nature* 337: 655–59
113. Richardson, J. S. 1981. *Adv. Protein Chem.* 34: 168–339
114. Ringwald, M., Schuh, R., Vestweber, D., Eistetter, H., Lottspeich, F., et al. 1987. *EMBO J.* 5: 3647–53
115. Roe, B. A., Ma, D.-P., Wilson, R. K., Wong, J. F.-H. 1985. *J. Biol. Chem.* 260: 9759 74
116. Roepe, P. D., Kaback, H. R. 1989. *Proc. Natl. Acad. Sci. USA* 86: 6087–91
117. Rothman, J. E., Kornberg, R. D. 1986. *Nature* 322: 209–10
118. Runswick, M. J., Powell, S. J., Nyren, P., Walker, J. E. 1987. *EMBO J.* 6: 1367–73
119. Schägger, H., Borchart, U., Aquila, H., Link, T. A., von Jagow, G. 1985. *FEBS Lett.* 190: 89–94
120. Schägger, H., Borchart, U., Machleidt, W., Link, T. A., von Jagow, G. 1987. *FEBS Lett.* 219: 161–68
121. Schleyer, M., Neupert, W. 1985. *Cell* 43: 339–50
122. Schofield, P. R., Darlison, M. G., Fujita, N., Burt, D. R., Stephenson, F. A., et al. 1987. *Nature* 328: 221–27
123. Schulz, G. E., Schirmer, R. H. 1979. *Principles of Protein Structure.* Berlin: Springer Verlag. 314 pp.
124. Serrano, R., Kielland-Brandt, M. C., Fink, G. R. 1986. *Nature* 319: 689–93
125. Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., et al. 1986. *EMBO J.* 5: 2043–49
126. Shull, G. E., Schwartz, A., Lingrel, J. B. 1985. *Nature* 316: 691–95
127. Sigrist, H., Wenger, R. H., Kislig, E., Wuthrich, M. 1988. *Eur. J. Biochem.* 177: 125–33
128. Simpson, L., Neckelmann, N., de la Cruz, V. F., Simpson, A. M., Feagin, J. E., et al. 1987. *J. Biol. Chem.* 262: 6182–96
129. Singh, S., Lowe, D. G., Thorpe, D. S., Rodriguez, H., Kuang, W.-J., et al. 1988. *Nature* 334: 708–12
129a. Sober, H. A., ed. 1970. *Handbook of Biochemistry*, pp. C-10–C-23, C-62–C-64. Cleveland: The Chemical Rubber Co. 2nd ed.
130. Stayton, M. M., Brosio, P., Dunsmuir, P. 1987. *Plant. Mol. Biol.* 10: 127–37
131. Steppuhn, J., Rother, C., Hermans, J., Jansen, T., Salnikow, J., et al. 1987. *Mol. Gen. Genet.* 210: 171–77
132. Stoffel, W., Hillen, H., Giersiefen, H. 1989. *Proc. Natl. Acad. Sci.* USA 81: 5012–16
133. Takeshima, H., Nishimura, S., Matsumoto, T., Ishida, H., Kangawa, K., 1989. *Nature* 339: 439–45
134. Tamkun, J. W., DeSimone, D. W., Fonda, D., Patel, R. S., Buck, C., et al. 1987. *Cell* 46: 271–82
135. Tanabe, T., Takeshima, H., Mikami, A., Flockerzi, V., Takahashi, H., et al. 1987. *Nature* 328: 313–18
136. Thiede, M. A., Ozols, J., Strittmatter, P. 1986. *J. Biol. Chem.* 261: 13230–35
137. Trebst, A. 1986. *Z. Naturforsch.* 41c: 240–45
138. Velours, J., Esparzq, M., Hoppe, J., Sebald, W., Guerin, B. 1984. *EMBO J.* 3: 207–12
139. Vergères, G., Winterhalter, K. H., Richter, C. 1989. *Biochemistry* 28: 3650–65
140. Verma, A. K., Filoteo, A. G., Stanford, D. R., Wieben, E. D., Penniston, J. T., et al. 1988. *J. Biol. Chem.* 263: 14152–59
141. Verner, K., Schatz, G. 1988. *Science* 241: 1307–13
141a. von Heijne, G. 1988. *Biochim. Biophys. Acta* 947: 307–33
142. von Heijne, G. 1986. *FEBS Lett.* 198: 1–4
142a. von Heijne, G. 1986. *EMBO J.* 5: 3021–27
143. Walker, J. E., Saraste, M., Gay, N. J. 1984. *Biochim. Biophys. Acta* 768: 164–200
144. Wetlaufer, D. B. 1981. *Adv. Protein Chem.* 34: 61–92
145. Wickner, W. 1988. *Biochemistry* 27: 1081–86
145a. Widger, W. R., Cramer, W. A., Herrmann, R. G., Trebst, A. 1984. *Proc. Natl. Acad. Sci. USA* 81: 647–78
146. Wood, D., Darlison, M. G., Wilde, R. J., Guest, J. R. 1984. *Biochem. J.* 222: 519–34
147. Wrubel, W., Stochaj, U., Sonnewald, U., Theres, C., Ehring, R. 1989. *Biol. Chem. Hoppe-Seyler* 370: 636 (Abstr.)
148. Yamaguchi, M., Hatefi, Y., Trach, K.,

Hoch, J. A. 1988. *J. Biol. Chem.* 263: 2761–67
149. Yarden, Y., Escobedo, J. A., Kuang, W.-J., Yang-Feng, T. L., Daniel, T. O., *et al.* 1986. *Nature* 323: 226–32
150. Yochem, J., Weston, K., Greenwald, E. 1988. *Nature* 335: 547–50
151. Young, I. G., Rogers, B. L., Campbell, H. D., Jaworowski, A., Shaw, D. C. 1981. *Eur. J. Biochem.* 116: 165–70
152. Deleted in proof
153. Zimmermann, R., Sagstetter, M., Lewis, M. J., Pelham, H. R. B. 1988. *EMBO J.* 7: 2875–80
154. Zuber, H. 1986. In *Encyclopedia of Plant Physiologie, New Series*, Vol. 19, *Photosynthesis III, Photosynthetic Membranes and Light Harvesting Centers*, ed. L. A. Staehlin, C. J. Arntzen, pp. 238–51. Berlin: Springer

Annual Review of Biophysics and Biophysical Chemistry
Volume 19, 1990

CONTENTS

PREFATORY

Whither Biophysics?, *Gregorio Weber* 1

STRUCTURAL PRINCIPLES

Mechanisms of Long-Distance Electron Transfer in Proteins: Lessons from Photosynthetic Reaction Centers, *Steven G. Boxer* 267

Electrostatic Interactions in Macromolecules: Theory and Applications, *Kim A. Sharp and Barry Honig* 301

Zinc Finger Domains: Hypotheses and Current Knowledge, *Jeremy M. Berg* 405

Ion Distributions around DNA and Other Cylindrical Polyions: Theoretical Descriptions and Physical Implications, *Charles F. Anderson and M. Thomas Record, Jr.* 423

STRUCTURE AND FUNCTION

The Proton-Translocating ATPase of *Escherichia coli*, *A. E. Senior* 7

Gramicidin Channels and Pores, *B. A. Wallace* 127

The Structure and Function of the Aspartic Proteinases, *David R. Davies* 189

Mechanisms of Cytoplasmic Hemoglobin and Myoglobin Function, *Jonathan B. Wittenberg and Beatrice A. Wittenberg* 217

Spectroscopic Analysis of Genetically Modified Photosynthetic Reaction Centers, *William J. Coleman and Douglas C. Youvan* 333

On the Microassembly of Integral Membrane Proteins, *Jean-Luc Popot and Catherine de Vitry* 369

DYNAMICS

Electrical Modulation of Membrane Proteins: Enforced Conformational Oscillations and Biological Energy and Signal Transductions, *Tian Yow Tsong* 83

(*continued*) v

EMERGING TECHNIQUES

NMR Studies of Metabolism, *Sebastian Cerdan and Joachim Seelig* 43

Molecular Structure and Vaccine Design, *S. Vajda, R. Kataoka, C. DeLisi, H. Margalit, J. A. Berzofsky, and J. L. Cornette* 69

Crystallography of Biological Macromolecules at Ultra-Low Temperature, *Håkon Hope* 107

Calorimetrically Determined Dynamics of Complex Unfolding Transitions in Proteins, *E. Freire, W. W. van Osdol, O. L. Mayorga, and J. M. Sanchez-Ruiz* 159

Uniform and Selective Deuteration in Two-Dimensional NMR of Proteins, *David M. LeMaster* 243

INDEXES

Subject Index 467

Cumulative Index of Contributing Authors, Volumes 15–19 475

Cumulative Index of Chapter Titles, Volumes 15–19 477