Compilation of published signal sequences

Marion E.E. Watson

Department of Molecular Genetics, G.D.Searle and Co. Ltd., PO Box 53, Lane End Road, High Wycombe, Bucks, UK

Received 27 April 1984; Revised and Accepted 11 June 1984

INTRODUCTION

The process by which proteins are transported across membranes was presented as a detailed hypothesis, the 'signal hypothesis' in 1975 (1) and subsequently in revised forms (2,3). An essential feature is the presence of a polypeptide of 20-40 amino acids, predominantly hydrophobic, at the amino terminus of secreted proteins. This binds to the endoplasmic reticulum and transport is initiated before translation is complete. Despite some contradictory evidence (e.g. 4-6) the essentials of the signal hypothesis have been widely accepted and have been extended to include proteins which become integrated into membranes (3,7) and bacterial systems (8).

The current hypothesis requires that the signal peptide must be transient. However, a number of secreted and membrane-bound proteins have now been shown to have N-terminal sequences which resemble the transient peptides, but which are not cleaved. It has been shown that these sequences can function as signals for translocation into the rough endoplasmic reticulum (9,10). Therefore, they have been included in this compilation.

Proteins which are transferred across two membrane bilayers into mitochondria and chloroplasts have also now been found to have transient amino terminal regions (3,11,12). A major difference in the transport of these proteins is that it occurs post-translationally and it is clear that their signals are quite different.

The structure, function and processing of signal sequences have been reviewed (7,8,13,14). There has been considerable interest as to which features of signal sequences are essential for their function, notably their secondary structure (15), hydrophobicity (16-19) and the positions of positively charged residues (20,21). The site of cleavage of a signal from its mature protein is of particular interest and a common format for the cleavage site of both prokaryotic and eukaryotic signals has been proposed (22-24). The enzyme responsible for the cleavage step in *E. coli* and its nucleotide sequence have been identified (25). This protein is itself one of the exceptions to the signal hypothesis in that its signal is not cleaved. The use of cloned DNAs with altered signal sequences (e.g. 26-33) and of cells with mutations in their secretory apparatus (e.g. 34-37) will help answer many of the unsolved questions on the process of secretion.

The following is a compilation of published signal sequences. The proteins are grouped mainly into genera. Phage and plasmid encoded proteins are grouped with their hosts. They are classified as inner membrane, outer membrane, periplasmic or transmembrane proteins where appropriate. The first ten amino acid residues of the following protein sequences are also given. The end of the signal may not always represent an *in vivo* cleavage site. Where there is clear evidence to the contrary this is indicated at the junction (see abbreviations). Note that each protein complete with its signal peptide should be termed 'pre' protein. In many cases, e.g. the human prefibrinogens, the signal sequence is derived only from DNA sequence data and the actual initiator methionine of several present is unknown. For some proteins the exact end of the signal is unknown. It is inferred by comparison with well characterised junctions using the method of 'processing probabilities' described by von Heijne (24).

The author would welcome any information regarding errors or missing sequences.

SOURCE & PROTEIN	REFS.	SIGNAL	JUNCTION
<u>EUKARYOTIC SIGNALS</u>			
BABOON			
α_1 -antitrypsin	38	...ILLLAGLCCLLP6SLA	EDPQGDAAQK
BOVINE			
proparathyroid	39	MMSAKDMVKVMIVMLAICFLARSDG	KSVKKRAVSE
growth hormone	40	MMAAGPRTSLLLAFALLCLPWTQVVG	AFPAMSLSGL
oxytocin/neurophysin I	41,42	MAGSSSLACCLLGLLALTSA?CYIQNCPL6G	
vasopressin/ " II	43	MPDATLPACFLSLLAFTSA?CYFQNCPR6G	
pro ACTH- β lipotrophin	44,45	MPRLCSSRSAGALLLALLQASMEVRGWCLES?SQCCDLTTES	
glicentin/glucagon	46	MKSLYFVAGLFVMLVQGSWQ	RSLQNTTEKS
α -pituitary glycoproteins	47,48	MDYYRKYAAVILTILSLFLQILHS	FDPGEFTMQG
protachykinin	49	MKILVAVAVIFFISTQLSA?EEIGANDDFN	
proalbumin	50	MKWVTFISLLLLFSSAYS	RGVFRDTHK
prorenin	51	MRCLVVLLAVFALSQG	AEITRIPLYK
Low MW kininogen 1&II	52	MKLITILFLCSRLPLSLT?QESSQEIDCN	
AChR α -subunit	453	MEPRPLLLLLGLCSAGLVLG	SEHETRLVAK
CANINE			
trypsinogen 1	54	...PLLILAFLEXAAVA	TPTDDDDKIV
trypsinogen 2 & 3	54,55	MALLFLFLALLLAYVA FL 6	FPIDDDDKLV
chymotrypsinogen	54	...AFLILVxAFALxxVAFx	xxVPAIxPVx
chymotrypsinogen 2	56	MAFLWLLSCFALLGTAFG	CGVPAIQPVL
proelastase	54	...FxVxVLAFLLAYAx	FxxIPxxPIA
procarboxypeptidase A1	54	...LILVFGALLxAIY	QxAFVxxxxx
amylase	54	...FFLLLxVIGFxVA	QYAPHxxxxx
proinsulin	57	MALWMRLLPLLALLALWAPAPTRA	FVNQHLCGSH
HAMSTER			
proglucagon	58	MKNIIYIVAGFFCGAGQGSWQ	HSLODTEEKS
proinsulin	59	MTLWMRLLPLLTLLVLWEPNPANA	FVNQHLCGSH
HUMAN			
growth hormone	60,61	MATGSRTSLLAFGLLCLPWLQEGSA	FPTIPLSRLF
somatocrinin	62,63	MPLWVFFFVILTLSNS?SHCSPPPPLT	
α -gonadotrophin	64	MDYYRKYAAIFLVTLVSFLHVLHS	APDVQDCPEC
pro plasminogen activator	65	MDAMKRGLCCVLLCGAVFVSPSQEIHA?RFRRGARSYN	
placental lactogen	66,67,68	MPGSRTSLLAFALLCLPWLQEAGA MAA MAP	VQTVPLSRLF
relaxin	69	MPRLFLFHLLEFCLLLNQFSRAVAA	KWKDDVIKLC

SOURCE & PROTEIN	REFS.	SIGNAL	JUNCTION
proinsulin	70	MALWMRLPLALLALWGPDAAA	FVNQHLCGSH
insulin like growth fact.	71	(...?)MSSSHLFYLALCLLTFTSSATA	GPETLCGAEL
pancreatic peptide	72	MAAARLCLSLLLLSTCVALLLQPLLGAQG	APLEPVYPGD
progastrin	73,74	MQRLCVYVLIIFALALAAFSEA?	SWKPRSQQPD
α_1 antitrypsin	75,76	MPSSVSWGILLLAGLCCLVPVSLA	EDPQGDAQK
prorenin	77	MDGWRMPRWGLLLLWGSCTFG?	LPTDITTFKR
α -interferon A	78	MALTFALLVALLVLSCKSSCSVG	CDLPQTHSLG
B	78	MALTFYLMVALVLSYKSFSSLG	CDLPQTHSLG
C	78	MALSFSLLMAVLVLSYKSICSLG	CDLPQTHSLG
D	78	MASPFALLMVLVLECKSSCSLG	CDLPETHSLD
E	78	...LPLG	CDLPOAHSVG
F	78	MALSFSLLMAVLVLSYKSICSLG	CDLPQTHSLG
H	78	MALPFSLLMALVLSCKSSCSLG	CNLSQTHSLN
β -interferon	79	MTNKCLLQIALLLCFSTTALS	MSYNLLGFLQ
γ -interferon	80	MKYTSYILAFQLCIVLGSIG	CYCQDPYVKE
interleukin 2	81	MYRMQLLSICIALSLA?	LVTNSAPTSS
Ig heavy chain	*82	MEFGLSWFLVAILKGVC	EVOLLES6GG
Ig κ chain-101	*83	MDMRVLAQLLGLLLCFPGARC	DIQMTQSPSS
Ig κ chain-102	*83	MDMRVPAQLLGLLLWLPGARC	DIQMTQSPST
HLA-DS α -chain	*84	MILDKALMLALGALTVMSPCGG	EDIVADDHVA
HLA-DR α -chain	*85,86	MAISGVVPLGFFIIAVLMSAQESWA	IKEEHVIIQA
HLA-DR β -chain	*87	MVCLKLPGGSSLAALTVTLMVLSSRLAFA	GDTRPRFLEL
HLA-DR β -chains 3&4	*88	MVCLRLPGGSCMAVLVTLMVLSSPLALA	GDTRPRFLEY E
HLA-DR β -chain 5	*88	...GGIPGDLRVATVTLMAILSSSLAEG	RDSPEDFVYQ
HLA-CW3	*89	MRVMAPRTLILLLSGALALTETWAGSHS?	MRYFCTAVSR
HLA-A3	*90	MARGDQAVMAPRTLILLLSGALALTQWAGSHS?	MRYFFTSVSR
apolipoprotein A-1	91,92,93	MKAAVLTIAVLFTGSQA	RHFWQQDEPP
prothrombin	94	...QLPGCLALAAALCSLVHS?	QHVFLAPQQA
antithrombin III	95	MYSNVIGTVTSGKRKVYLLSLLIGFWDCVTC	HGSPVDICTA
α -fibrinogen	96,97	MFSMRIVCLVLSVVGTAWT	ADSGEGDFLA
β -fibrinogen	98,99	MKRMVSWSFHKLKTMKHLLLLVCVFLVKS	QGVNDNEEGF
γ -fibrinogen	99	MSWSLHPRNLLIFYFYALLFLSSTCVA	YVATRDNCCI
α -haptoglobin	100,101	MSALGAVIALLLWGQLFA	VDSGNDVTDI
prosomatostatin	102	MLSCRLOCALAALSIVLALGCVTGA?	PSDPRLRQFL
proenkephalin A	103,104	MARFLTICTWLLLGPGLLA?	TVRAECSQDC
proenkephalin B	105	MAWQGLVLAACLLMFPSTTA?	DCLSRCSLCA

SOURCE & PROTEIN	REFS.	SIGNAL	JUNCTION
pro ACTH- β lipotrophin	106	MPRSCCSRSYGALLLALLLQASMEVRG	WCLESSQCQD
α -fetoprotein	107	MKWVESIFLIFLLNFTESR	TLHRNEYGIA
retinol binding protein	108	MKWVWALLLLAAGAAA	ERDCRVSSFR
AChR α -subunit	*53	MEPWPLLLLFSLCSAGLVLG	SEHETRLVAK
MONKEY			
proinsulin	109	MALWMRLPLALLALWGPDPA	FVNQHLCGSH
MURINE			
Ig heavy chain-315	*110,111	MKVLSELLYLTAIPHIMS	DVQLQESGPG
Ig λ_1 chain-104E	*112,113,114	MAWISLILSLALSSGAIS	QAVVTQESAL
Ig λ_2 chain-351	*113,114	MAWTSILISLALCSGASS	QAVVTQESAL
Ig λ chain-315	*111,113	MAWTSILISLIALCSGAS	QAVVTESALT
Ig κ chain-41A	*112,114	MDMRAPAQIFGFLLLFPGTRC	DIQMTQSPSS
Ig κ chain-41B	*112	MRAPAQIFGFLLLFPGTRC	DIQMTQSPSS
Ig κ chain-321 & 63	*114,115	METDTLLWVLLWVPGSTG	DIVLTQSPAS N
Ig κ chain-2	*116	MSVLTQVLALLLWLTGARC	DIQMTQSPAS
Ig v-heavy chain	*117,118	MKLWLNWVFLTLHGINC	EVKLVESGGG
HLA H-2kd	*119,120	MAPCTLLLLAAALAPTQTRA	GPHSRLRYFT
HLA I-E α d	*121	MATIGALVLRFFFIIVLMSQKSWA	IKEENTIIOA
HLA I-E β d	*122	MVWLPRVPCVAAVILLTVLSPPVAL	VRDTRPRFLE
epidermal growth factor	123	MPWGRRPWLLAFLLVFLKISILSVTA	?WQTGNCQPGP
mast cell growth factor	124	MVLASSTTSIHTMLLLMLFHLGLQA	?SISGRDTHRL
pro ACTH- β lipotrophin	125,126	MPRFCYSRSYGALLLALLLQTSIDVWS	WCLESSQCQD
α_1 -interferon	127	MARLCAFLMVLAVMSYWPTCSLG	CDLPQTHNLR
α_1 -interferon	127	MARLCAFLVLMIVMSYWSICSLG	CDLPHTYNLR
β -interferon	128	MNNRWILHAAFLLCFSTTALS	INYKQLQLQE
γ -interferon	129	MNATHCILAQLFLMAVSG	CYCHGTVIES
interleukin 3	130	MVLASSTTSIHTMLLLMLFHLGLQA	?SISGRDTHRL
α -fetoprotein	131	MKWITPASLILLHFAASK	ALHENEFGIA
complement component 3	132	MGPASGSQLLVLLLLASSPLALG	IPMYSIITPN
prorenin	133	MDRRRMPLWALLWSPCTFSLPTGTT	?FERIPLKKMP
submaxillary gland pr. 3	134	MKFLALLVLLGVSTILVS	?CQDPETNSTE
OVINE			
β -caesin	135,136	MKVLILACLVALALA	REQEELNVVG
αS_1 -caesin	135,136	MKLLILTCLVAVALA	RPKHPIKHxG
αS_2 -caesin	136	MKVLMAKCLVAVALA	KxxMxxVxxx
αK -caesin	135,136	MRKSILLVVTILALTPLFLIA	QEQNQEQRIC

SOURCE & PROTEIN	REFS.	SIGNAL	JUNCTION
β -lactoglobulin	136	MKCLLLALGLALACGVQA	IIVTQTMKGL
α -lactalbumin	137	MMSFVSLLLVGILFWATQA	EQLTKCEVFO
ACTH releasing factor	138	MRLPLLVSQVGLLVA?	LLPSPPCRAL
tropoelastin	139	MKxLTAAARRPGVLFLLLCILFCSP	GGVPGAVP6G L P
PORCINE			
growth hormone	140	...TSVLLAFALLCLPWTQEVG	AFPAMPLSSL
proenkephalin B	141	MAWQGLLLAACLLVLPSTMA?	DCISGCSLCA
pro ACTH- β lipotrophin	142	MPRLCGSRSGALLTLLLQASMVGVRG	CLESSQCQDL
relaxin	69	MPRLFSYLLGVWLLLSQLPREIPG	QSTNDFIKAC
progastrin	143	MQRLCAYVLIHVLALAACSEA?	SWKPGFQLQD
α -lactalbumin	144	MMSFVSLLVVGILFPAIQA?	KQFTKCELSQ
RABBIT			
IgG VH α 2del2,15	*82	METGLRWLLLVAVLKGVQC	QSVKESEGGL
Ig κ chain-b5	*145	...TRAPAQLLGLLLWLPGARCA	DVMTQTPAS
uteroglobin	146	MKLAIFLAVxLALLGSPASA	GICPRFAHVI
cytochrome P-450	*147	MEFSLLLLLAFLAGLLLLLFRG/HPxAxGxLxx	
α -lactalbumin	144	MMPLVPLLLVSIVFPGIOA?	TOLTRCELTE
RAT			
trypsin I	148	MSALLILALVGAAVA	FPLEDDDKIV
trypsin II	148	...LVGAAVA	FPVDDDDKIV
α_1 -antitrypsin	149	MAPxxxxxLLLLAxLCCLAPxxLA	DAxxxDxxxx
elastase I	148,150	MLRFLVFASLVLYGHS	TQDFPETNAR
elastase II	148,150	MIRTLLSAFVAGALS	C6YPTYEVQH
chymotrypsin	148	MAFLWLVSFCFALVGATFG	Cxxxxxxxxx
procarboxypeptidase A	148,151	MKRLILSLLEAVCG	NENFVGHQVL
amylase	148,152	MKFVLLLSLIGFCWA	QYDPHTADGR
proinsulin	153,154,155	MALWMRFLPLLALLVLWEPKPAQA	FVKQHLGCPH I I R
ribonuclease A	148,156	MGLEKSLFSLSLVLVLGWQPSLG	GESRESSADK
growth hormone	157	MAADSQTPLLTLTSLCLLWPQEA	LPAMPLSSLF
angiotensinogen	158	MTPTGAGLKATIFCILTWSLTAG	DRVYIHPFHL
oxytocin/neurophysin I	159	MACPSLACCLLGLLALTS?	CVIONCPLGG
vasopressin/ " II	160	MLAMMLNTTSLACFLSLALTS?	CYFQNCPRGG
relaxin	161	MSSRLLLQLLGFWFLSQPCRA	RVSEEWMDQV
prosomatostatin	162	MLSCRLQCALAALCIVLALGGVTGA?	PSDPRLRQFL
prostatein C1	163	MSTVELSLCLLIMLAVCCYEANA?	SOICELVAHE
prostatein C2	163	MRLSLCLLTILVVCYEANG?	QTLAGVCQAL

SOURCE & PROTEIN	REFS.	SIGNAL	JUNCTION
prostatein C3	164	MKLVLFLLVITIPICCYA	SGS6CSILDE
proalbumin	165,166	MKWVTFLLLLFISGSASF	RGVFRREAHK
procholecystokin	167	MKCGVCLCVVMAVLAAGALA?	QPVPVPEAVD
prolactin	168	MNSQVSARKAGTLLLLMMSNLLFCQNVQT	LPVCS6GDCQ
whey phosphoprotein	169	MRCFISLVGLLLEVALA	RNLQEHVFN
uteroglobulin	170	MKLAITLALVTIALLCSPASA	GICPRFAXVI
haptoglobin	171	MxALxAVIxLLLxQFA	ADFGNEVTDI
α_1 -acid glycoprotein	172	MALHMLVVLVLSLLPLLEA	QNPEPANITL
α_1 -interferon	173	MARLCAPLMSLVVVSYSACCLG	CDLPHTHNL
apolipoprotein A	174	MKAVALVALVFLTGCA	xEFxQQDEPQ
apolipoprotein E	175	MKALWALLVPLLTGCLA	EGELEVTDL
seminal vesicle pr. IV	176	MKSTSLFLCSLLLLLVGTGAIG	RKTKxxxxxx
procalcitonin	177	MGFLKFSFPLVVSILLYQACGLQA?	VPLRSTLESS
lutropin	178	MERLQGLLLWLLSPSVVWA	SRGPLRPLCR
cytochrome P-450	*179,180	MEPSILLLLALLVGFLLLLVRG/	xxxxxxxxxx T
cytochrome P-450a	*180	MLDTGLLLVVILATLTVMLLLLT/	xxxxxxxxxx
CHICKEN			
proinsulin	57,181	MALWIRSLPLLALLVFS6PGTSYA	AANQHLC6SH
lysozyme	*182	MRSLLILVLCFLPLAALG	KVFGRCELAA
ovomucoid	183	MAMAGVFVLFVLCGFLPDAAF	AEVDCSxxFP
conalbumin	183	MKLILCTVLSLGIAAVCF	APPKSVIRWC
vitellogenin	184	MRGIIALVLTIVGS?	QKFDIDPGFN
pro $\alpha 2(1)$ collagen	185	MLSFDVTRILLLLA?	VTSYLATSQH
pro $\alpha 1(1)$ collagen	185	MFSFVxSRLLLIA?	AxxLLxxxxx
Apo VLDL-II	186	MQYRALVIAVILLSTTVPEVCS	KSIIDRERRD
Ig κ -chain	187	...SGSLVQA	ALTQASVSA
tropoelastin b	188	MRQAAAPLLPGVLLLSILPASQ	GGVPGAIPGG
CAIMAN CROCODYLUS			
Ig(v)heavy chain	*118	MGLGLHLLVLAAALQGAWS	QVQLVESG6D
XENOPUS			
neurotensin like peptide	189	MYKGIFLCVLLAVICANSLA?	TPSSDAEDN
ANGLER FISH			
proinsulin	155,190	MAALWLQSFSLVLLVVSWP6SQA	VAPAxLxxx
proglucagon	59,191	MKRIHSLAGILLVLGLIQSSC?	RVLMOEADFS
prosomatostatin	102	MKMVSSSRLRCLLVLLSLTASISCSFA?	GNRDSKLRLL

SOURCE & PROTEIN	REFS.	SIGNAL	JUNCTION
CARP			
proinsulin	190	MAVWIQAGALLFLAVSSVNA	MAGAPQHLGG
CATFISH			
prosomatostatin	192	MSSSPRLRLALALMCLVSAVGVIS?	CGRPHVVLNS
HAGFISH			
proinsulin	155,190	MALSPFLAAVIPLVLLSRAPPSADT	RTTGxxxxxx
SALMON			
proinsulin	155	MALWLQAASLLVLLALSPGVDA	AAQHLGGSH
TORPEDO			
AChR α -subunit	193,194	MILCSYWHVGLVLLFSCCGLVLG	SEHETRLVAN
WINTER FLOUNDER			
antifreeze protein I	195	MRITEANPDPDAKAVPAAAAP	DTASDAAAAA
antifreeze protein II	196	MALSLFTVGLLIFLFWT?	MRITEASDPD
BEE			
promelittin	197,198	MKFLVNVALVFMVYISYIYA	APEPEPAPEP
DROSOPHILA MELANOGASTER			
68C glue protein-8	199	MKLLVVAVIACIMLIGFADPASG	CKDCSCVICG
68C glue protein-7	199	MKLIAVTIIACILLIGFSDLALG	GACEGQPCGP
68C glue protein-3	199	MKLTIALALASILLIGSANVANC	CDCGCPTTTT
yolk protein 1	200	MNPMRVLSSLACLAVAL?	ALAKPNGRMD
BRASICA NAPUS			
napsin	201	MANKLFLVSATLAFFFLLTNA?	SIYRTVVEFD
HORDEUM VULGARE			
α -amylase	202	MGKNGSLCCFSLLLLLLLAGLAS?	GHQVLFQGFN
PHASEOLUS VULGARIS			
phaseolin	203	MMRARVPLLLLGILFLASLSASFA?	TSLREEEESQ
PISUM SATIVUM			
vicilin	204	MLLAIAFLASVCVSS	RSDQENPFIF
prolectin	205	MASLETEMISFYAIFLSILLTTILFFKVNS	TETTSFLITK
ZEA MAIZE			
zein protein 19	206,207	MAAKIFCLIMLLGLSASAATA	SIFPQCSQAP T
zein protein 22.1	208	MATKILALLALLLVSATNA	FIIPQCSLAP
zein protein 22.3	208	MATKILSLLALLALFASATNA	SIIPQCSLAP
SACCHAROMYCES			
pro α factor 1	9,209	MRFPSIFTAVLFAASSALA/APVNTTTEDE	
pro α factor 2	210	MKFISTFLTFILAAVSVTAS?	SDEDIAQVPA

SOURCE & PROTEIN	REFS.	SIGNAL	JUNCTION
acid phosphatase	+211	MFKSVVYSILAASLANA	GTIPLGKLAD
invertase	+212	MLLQAFLLLAGFAAKISA	SMTNETSDRP
PLASMODIUM KNOWLESII			
circumsporozoite	*213	MKNFILLAVSSILLVDLLPT?	HFEHNVDLSP
TRYPANOSOMA BRUCEI			
surface glycoprotein 117	*214	(...?)MLTSLLYAITPADG	AKEALEYKTW
surface glycoprotein 221	*214	(...?)MPSNQEARLFLAVLVLAQVLPILVDS	AAEKGFQKAF
VIRAL SIGNALS			
ADENO VIRUS 2			
19K glycoprotein	*215	MRYMILGLLALAAVCSAA	KKVEFKEPAC
HERPES SIMPLEX			
glycoprotein D-1	216,217	MGGTAARLGAVILFVVIVGLHGVRG	KYALADASLK
glycoprotein D-2	218	MGRLTSGVGTAAALLVVAVGLRVVCA	KYALADPSLK
AVIAN INFLUENZA			
A/Rostock HA	*219	MNTQILVFALVAVIPTNA	DKICLGHHAV
HUMAN INFLUENZA			
A/Victoria HA	*220	MKTIIALSIFYCLVFA	QDLP6NDNNS
A/Japan HA	*221	MAIIYLILLFTAVRG	DQICIGYHAN
A/WSN HA	*222	MKAKLLVLLYAFVAG	DQICIGYHAN
B/Singapore HA	*223	MKAIIVLLMVVTSNA	DRICTGITSS
B/Lee HA	*224	MKAIIVLLMVVTSNA	DRICTGIT6S
A/Udorn NA	*225	MMNPNQKIITIGSVSLTIA/TIGFLMQIAI	
A/PR8 NA	*226	MNPNQKIITIGSICLVGLIS/LILQIGNIIS	
A/WSN NA	*10,227	MNPNQKIITIGSICMVV6IIS/LILQIGNIIS	
B/Lee NA	*228	MLPSTVQTLTLLTSG6VLLSLYVSAS/LSYLLYSDVL	
MOUSE MAMMARY TUMOUR VIRUS			
envelope glycoprotein	*229	MNLAPIKEXTAWQHLOALISEAEVLTSTQPN5c	LTFLALLSVL6PPPVTG ESYWAYLPKP
RABIES VIRUS			
CVS glycoprotein	*230	MVPQVLLFVLLGFSLCFG	KFPIYTIPDK
ROUS SARCOMA VIRUS			
envelope glycoprotein	*231	MRRALFLOAFTGYP6KTSKKDSKEKPLATSKKDc	PEKTPLLPTRVNYILIGVLVLC EVTGVRA DVHLL EQPN
SIMAN ROTAVIRUS-SA11			
glycoprotein NCVP5	*232	MEKLTDLNYTSLSVITLMNNTLHT?	ILEDPGMAYF
VESICULAR STOMATITIS VIRUS - HAMSTER			
glycoprotein G	*233	MKCLLYLAFLFIHVNC	KFxIVFPxxx

SOURCE & PROTEIN	REFS.	SIGNAL	JUNCTION
VESICULAR STOMATITIS VIRUS - HUMAN			
N.J.Ogden glycoprotein	#234	MLSYLIFALAVSPILG	KIEIVFPQHT
Ind.SanJuan glycoprotein	#234	MKCLLYLAFLFIGVNC	KFTIVFPHNQ
Cocal glycoprotein	#235	MKFLLLLFixLPLxxx	KFSIVFPQNO
N.J.Concan glycoprotein	#235	MLxPLIFxLxVxPILx	KIEIVFPQHx
Ind.Toronto glycoprotein	#235	MKCLLYLxFLFPxVxG	KFTIVFPYNQ
YEAST KILLER			
M1 protoxin	236,237	MTKPTQVLVRSVSILFFITLLHLVVA?	LNDVAGPAET
<u>PROKARYOTIC SIGNALS</u>			
BACILLUS AMYLOLIQUIFACIENS			
α -amylase	238	MIQKRKRTVSFRLVLMCTLLFVSLPITKTS	VNGTLMQYFE
prosubtilisin	239	MRGKKVWISLLFALALIFTMAFGSTS?	SAQAAGKSNG
BACILLUS CEREUS			
P-lactamase	+240,241	MMILKNKRMLKIGICVGILGLSITSLEAFc TGESLQVEAKEKTGV	KHKNQATHKE
BACILLUS LICHENIFORMIS			
P-lactamase	+242	MKLWFSTLKLKAAAVLLFSCVALAGCANNQNA	SQPAEKNEKT
α -amylase	243	MKQQKRLYARLLTLFALIFLLPHSAAAA	ANLNGTLMQY
BACILLUS SUBTILIS			
α -amylase	244,245,246	MFAKRFKTSLLPLFAGFLLLFHLVLAc Y	GPAAASA ETANKSNELT
BACTEROIDES NODOSUS			
pilin	#247	...FTLIELMIVVAIIIGILAAFAIPA/YNDYIAR	SQA
CORYNEBACTERIUM DITHERIAE			
phage proDiphtheria toxin	248,249	MSRKLFASILIGALLGIGAPPSAHA	GADDVVDSSK
ESCHERICHIA COLI			
α -amylase	244	MFAKRFKTSLLPLFAGFLLLFHLVLAGPAAAS	AETANKSNEL
h.l.t.-B (human)	250	MNKVKCYVLFTALLSSLCAYLFTALLSSLYAHG	APQSITELxx
h.l.t.-B (porcine)	251	MNKVKCYVLFTALLSSLYAHG	APQTITELCS
h.l.t.-A (porcine)	252	MKNITFIFFILLASPLYA	NGDRLYRADs
h.s.t. I	253	MKKLMLAIFISVLSFSPFSQS?	TESLDSSKEK
h.s.t. II	254,255	MKKNIAFLASMFVFSIATNAYA?	STQSNKKDLC
lipoprotein	#256	MKATKLVLGAVILGSTLLAG	CSSNAKIDEL
serine receptor	#257	MLKRIKIVTSLLLVL	A?VFGLLQLTSG
λ -receptor	#258,259	MMITLRKLPLAVAVAAGVMSAQAMA	VDFHGYARSG
ompA	#260,261	MKKTAIAlAVALAGFATVAQA	APKDNTWYTG

SOURCE & PROTEIN	REFS.	SIGNAL	JUNCTION
ompC	#262	MKVKVLSSLVPALLVAGAANA	AEVYNKDGNK
omp E(Ic)	#263	MKKSTLALVVMGIVASASVQA	AEIYNKDGNK
ompF	#264	MMKRNLAVIVPALLVAGTANA	AEIxxxxxxx
tolC	#265	MKKLLPILIGLSLSGFSLSQA	ENLMQVYQQA
pap pili subunit	#266	MIKSVIAGAVAMAVVSVGVNNA	APTIPQGQ6K
β -lactamase -chromosomal	+267	MFKTTLCALLITASCSTFA?APQQINDIVH	
β -lactamase -plasmid	+268,269	MSIQHFRVALIPFFAAFCPLPVFA	HPETLVKVKD
tonB	+270	MIMTSMTLDLPRFPWPTLLSVCIHGAVVAGLL?YTSVHQVIEL	
alkaline phosphatase	+271,272	MKQSTIALALLPLLFTPTVTKA	RTPEMPVLEN
phosphate binding pr.	+273	MKVMRTTVATVVAATLSMSAFSVFA	EASLTGAGAT
maltose binding protein	+274	MKIKTGARILALSALTTMFSASALA	KIEEGKLVIV
leu-specific binding pr.	+275	MKANAKTIIAGMIALAISHTAMA	DDIKVAVVGA
leu-ile-val binding pr.	+264	MNIKGKALLAGCIALAFSNMALA	xxxxxxxxxx
arabinose binding pr.	+276	MKxTKLVLGAVILTAGLSxGAXA	ENLKLGLxx
lactose permease	-277	MYYLKNTNFWMFGLFFFFFYFFIMGA/YFPFFPIWLH	
NADH dehydrogenase	-278	MTTPLKKIVIVGGGAGGLEMAT/QLGHKLGRKK	
phage M13 major coat pr.	-279,280	MKKSLVLKASAVATLVPMLSFA	AEGDDPAKAA
phage M13 minor coat pr.	-280	MKKLLFAIPLVVPFYSHS	AETVESCLAK
leader peptidase	-281	MANMFALILVIATLVGT/ILWCVDKFFF	
ERWINIA AMYLOVORA			
lipoprotein	#282	MNRTKVLVLGAVILGSTLLAG	CSSNAKIDQL
ENTEROBACTER AEROGENES			
ompA	#283	MKKTAAIAAVALAGFATVAGA	APKDNTWYAG
HALOBACTERIUM HALOBIUM			
bacteriorhodopsin	#284	MLELLPTAVEGV	QAQITGRPEW
M. MORGANII			
lipoprotein	#285	MGRSKIVLGAVVLASALLAG	CSxxxxxxxx
MORAXELLA NONLIQUEFACIENS			
pilin	#247,286	...FTLIELMIVIAIIGILA/AIALPAYQDY	
NEISSERIA GONORRHOEA			
pilin	#247,287	...FTLIELMIVIAIVGILA/AVALPAYQDY	
PSEUDOMONAS SP.			
pilin	#288	...FTLIELMIVVAIIGILA/AIAIPOYQNY	
carboxypeptidase G2	289	MRPSIHRTAIAAVLATAFVAGT	ALAQRDNVL
SALMONELLA TYPHIMURIUM			
ompA	#290	MKKTAAIAAVALAGFATVAGA	APKDNTWYAG

SOURCE & PROTEIN	REFS.	SIGNAL	JUNCTION
his binding protein	-291,292	MKKLALSLSLVAFSSATAAFA	AIPQKIRIGT
lys-arg-orn binding pr.	-291	MKKTVLALSLLIGLGATAASYA	ALPQTVRIGT
SERRATIA MARCESCENS			
lipoprotein	#293	MNRTKLVLGAVILGSHSLAG	CSSNAKIDQL
SHIGELLA DYSENTERIAE			
ompA	#294	MKKTATAITVALAGFATVAQA	APKDNNTWYT
STAPHYLOCOCCUS AUREUS			
staphylokinase	295	MLKRSLLFLTLLLLFSFS?	SITNEVSASS
β -lactamase	#296	MKKLIFLIVIALVLSACNSNSSHA	KELNDLEKKY
VIBRIO CHOLERA			
toxin subunit A	297	MVKIIFVFFIFLSSFSYA	NDDKLYRADS
toxin subunit B	297,298	MIKCLKFGVFFTVLLSSAYAHG	TPQNITDLCA
MITOCHONDRIAL SIGNALS			
NEUROSPORA CRASSA			
ATP synthase	299	MASTRVLASRLASQMAASAKVARPAVRVAQVSKRc TIQTGSPLOTLKRTQMTSIVNATTRAQAFQKRA	TSSEIAQAMV
SACCHAROMYCES			
cytochrome-C peroxidase	300	MTTAVRLLPSLGRTAHKRSLYLFSAAAAAc AAAAATFAYSQSHKRSSSSPGGGSNHGc WNNWGKAAALAS	TTPLVHVASV
EF-Tu (TufM)	301	MSALLPRLLTRTAFKASGKLLRLSSVISRc TFSQTITS?	YAAAFDRSKP
CHLOROPLAST SIGNALS			
CHLAMYDOMONAS REINHARDTII			
RuBPCase	302	MAVIAKSSVSAAVARPARSSVRPMAALKPAVKc S	AAPVAAPAEADD MMVWTPVNNK V
PISUM SATIVUM			
RuBPCase	303	...GPFGLKFMTGFPVKKVNTDITSic TSNGGRVKC	MQVWPPIGKK
SOYBEAN			
RuBPCase	304	MASSMISSPAVTTVNRAAGMVPFTGLKSc MAGFPTRKTNDITSIASNGGRVQC	MQVWPPIGKK
TRITICUM AESTIVUM			
RuBPCase TS234	305	...VAPFQGLKSTAGLPVSRSSNGASLc GSVSNNGGRIRC	MQVWPIEGIK
RuBPCase WS4.3	305	MAPAVMASSATTVAPFQGLKSTAGLPVSRRC SRGSLGSVSNNGGRIRC	MQVWPIExxx
RuBPCase W9	306	MAPAVMASSATTVAPFQGLKSTAGLPISCRSc GSTGLSSVSNNGGRIRC	MQVWPIExxx

ABBREVIATIONS:

... = incomplete sequence data; * = eukaryotic-membrane protein; # = bacterial outer membrane protein; - = bacterial inner membrane protein; + = periplasmic protein; @ = lysosomal protein; ■ = transmembrane protein; ? = proposed end of signal; / = N-terminus not cleaved; AChR = acetyl choline receptor; ACTH = corticotrophin; Ag = antigen; c = continued; HA = haemagglutinin; h.l.t. = heat labile toxin; h.s.t. = heat stable toxin; Ind. = Indiana; MW = molecular weight; NA = neuraminidase; N.J. = New Jersey; orn = ornithine; pr. = protein; RuBPCase = ribulose-1,5-bisphosphate carboxylase (small subunit); x = unknown amino acid; A = ala; C = cys; D = asp; E = glu; F = phe; G = gly; H = his; I = ile; K = lys; L = leu; M = met; N = asn; P = pro; Q = gln; R = arg; S = ser; T = thr; V = val; W = trp; Y = tyr.

DISCUSSION

A total of 277 sequences are presented here. A detailed analysis of their structure is beyond the scope of this report. However, their most striking similarities include their length, their hydrophobicity and the limited number of amino acids which occur at the carboxyl termini of the signals. The general format of a signal seems to include a charged residue within the first five amino acids, followed by a core of at least nine hydrophobic residues which should be sufficient to span a membrane. A helix breaking residue (glycine or proline) or a large polar residue (notably glutamine), frequently occurs four to eight residues before the cleavage site. The pattern of amino acids near the cleavage sites of 78 eukaryotic and viral proteins has been discussed in detail by von Heijne and a scheme suggested for calculating the probability of processing after each residue (24). The most striking feature of cleavage sites is the presence of amino acids with small, uncharged side chains at the carboxyl termini of the signals. If only those where the cleavage site is well characterised are considered, of 136 of the eukaryotic signals, 69% end with alanine or glycine and a further 24% end with serine, cysteine or threonine. If 16 viral signals are included, these percentages become 71% and 22% respectively. For the 40 well characterised prokaryotic signals 88% end with alanine or glycine and a further 10% with serine, cysteine or threonine.

Processing probabilities have been calculated for each sequence in this compilation (data not shown). For 74% of the eukaryotic and viral proteins the observed cleavage site has the highest processing probability. For a further 11% the site with the second highest probability is cleaved. These probabilities are based on data from eukaryotic and viral proteins only. However, for comparison, they were also calculated for the prokaryotic sequences. 39% of the observed sites were in complete agreement with the predictions. A further 15% are cleaved at the second most probable site. The process by which prokaryotic proteins are exported is more complex than for eukaryotic proteins since the correct targeting of proteins to the inner or outer membranes or to the periplasmic space must occur. Information in the mature protein sequences may be responsible for their ultimate localisation (307-310). Experimental evidence for the mechanisms involved is as yet insufficient and in some cases is contradictory (30,311).

Some of the signal sequences are unusually long, notably those of the Bacilli (31-44 residues) and the viral envelope glycoproteins (53 and 64 residues). In the latter the most hydrophobic regions are towards the carboxyl termini of their signals so it may be that these represent 'internal' signals. The signals responsible for targeting proteins to mitochondria and chloroplasts are also long (42-68 residues). They are very different to the other signals presented. This is not surprising since they must be clearly distinguished from 'export' signals and because their transport is initiated only after translation is complete. It is not possible to describe a general structure from the small sample available.

BIBLIOGRAPHY

The references are presented in the abbreviated form first author, year, journal, volume and page numbers.

1. Blobel G. (1975), J. Cell Biol., 67, 835-851.
2. Blobel G. (1979), Symp. Soc. Exp. Biol., 33, 9-39.
3. Blobel G. (1980), P.N.A.S. USA, 77, 1496-1500.
4. Morlon J. (1983), J. Mol Biol., 170, 271-275.
5. Nakamura M. (1983), Mol. Gen. Genet., 191, 1-9.
6. Nathans J. (1983), Cell, 34, 807-814.
7. Wickner W. (1980), Science, 210, 861-868.
8. Michaelis S. (1982), Annu. Rev. Microbiol, 36, 345-365.

9. Bos T.J. (1984), P.N.A.S. USA, 81, 2327-2331.
10. Julius D. (1984), Cell, 36, 309-318.
11. Neupert w. (1981), Trends Biochem. Sci., 6, 1-4.
12. Schatz G. (1983), Cell, 32, 316-318.
13. Kreil G. (1981), Ann. Rev. Biochem., 50, 317-348.
14. Suominen I. (1983), Int. J. Biochem. 15, 591-601.
15. Austen B.M. (1979), FEBS lett., 103, 308-313.
16. von Heijne G. (1981), Eur. J. Biochem., 116, 419-422.
17. von Heijne G. (1982), J. Mol. Biol, 159, 537-541.
18. Emr S.D. (1983), P.N.A.S. USA, 80, 4599-4603.
19. Finkelstein A. (1983), FEBS lett., 161, 176-179.
20. Inoye S. (1982), P.N.A.S. USA, 79, 3438-3441.
21. Vlasuk G.P. (1983), J. Biol. Chem, 258, 7141-7148.
22. Perlman D. (1983), J. Mol. Biol., 167, 391-409.
23. von Heijne G. (1984), J. Mol. Biol., 173, 243-251.
24. von Heijne G. (1983), Eur. J. Biochem., 133, 17-21.
25. Wolfe P.B. (1983), J. Biol. Chem, 258, 12073-12080.
26. Bedouelle H. (1980), Nature, 285, 78-81.
27. Emr S.D. (1980), Nature, 285, 82-85.
28. Hortin G. (1981), Cell, 24, 453-461.
29. Boeke J.D. (1980), J. Mol. Biol., 144, 103-116.
30. Koshland D. (1982), Cell, 30, 903-914.
31. Hall M.N. (1983), EMBO J., 2, 15-19.
32. Michaelis S. (1983), J. Bacteriol., 154, 366-374.
33. Inouye S. (1983), EMBO J., 2, 87-91.
34. Novick P. (1981), Cell, 25, 461-469.
35. Ito K. (1983), Cell, 32, 789-797.
36. Kumamoto C.A. (1983), J. Bacteriol, 154, 253-260.
37. Shiba K. (1984), EMBO J., 3, 631-635.
38. Kurachi K. (1981) P.N.A.S. USA, 78, 6826-6830.
39. Habener J.F. (1978), P.N.A.S. USA, 75, 2616-2620.
40. Woychik R.P. (1982), Nucleic Acids Res., 10, 71-97.
41. Land H. (1983), Nature, 302, 342-344.
42. Ruppert S. (1984), Nature, 305, 554-557.
43. Land H. (1982), Nature, 295, 299-303.
44. Nakanishi S. (1981), Eur. J. Biochem. 115, 429-438.
45. Nakanishi S. (1979), Nature, 278, 423-427.
46. Lopez L.C. (1983), P.N.A.S. USA, 80, 5485-5498.
47. Edwin C.R. (1983), Biochemistry, 22, 4856-4860.
48. Goodwin R.G. (1983), N.A.R., 11, 6873-6882.
49. Nawa H. (1983), Nature, 306, 32-36.
50. MacGillivray R.T.A. (1979), Eur.J.Biochem., 98, 477-485.
51. Moir D. (1982), Gene, 19, 127-128.
52. Nawa H. (1983), P.N.A.S. USA, 80, 90-94.
53. Noda M. (1983), Nature, 305, 818-823.
54. Carne T. (1982), J. Biol. Chem., 257, 4133-4140.
55. Devillers-Thiery (1975), P.N.A.S. USA, 72, 5016-5020.
56. Pinsky S.D. (1983), P.N.A.S. USA, 80, 7486-7490.
57. Kwok S.C.M. (1983), J. Biol. Chem., 258, 2357-2363.
58. Bell G.I. (1983), Nature, 302, 716-718.
59. Bell G.I. (1984), Diabetes, 33, 297-300.
60. Martial J.A. (1979), Science, 205, 602-607.
61. Denoto F.M. (1981), N.A.R., 9, 3719-3730.
62. Gubler U. (1983), P.N.A.S. USA, 80, 4311-4314.
63. Mayo K.E. (1983), Nature, 306, 86-88.
64. Fiddes J.C. (1979), Nature, 281, 351-356.

65. Pennica D. (1983), *Nature*, 301, 214-221.
66. Sherwood L.M. (1979), *P.N.A.S. USA*, 76, 3819-3823.
67. Barrera-Saldana H.A. (1983), *J. Biol. Chem.*, 258, 3787-3793.
68. Seeburg P.H. (1982), *DNA*, 1, 239-249.
69. Hudson P. (1983), *Nature*, 301, 628-631.
70. Bell G.I. (1979), *Nature*, 282, 525-527.
71. Jansen M. (1983), *Nature*, 306, 609-611.
72. Boel E. (1984), *EMBO J.*, 3, 909-912.
73. Boel E. (1980), *P.N.A.S. USA*, 80, 2866-2869.
74. Kato K. (1983), *N.A.R.*, 11, 8197-8203.
75. Leicht M. (1982), *Nature*, 297, 655-659.
76. Bollen A. (1983), *DNA*, 2, 255-264.
77. Imai T. (1983), *P.N.A.S. USA*, 80, 7405-7409.
78. Goeddel D.V. (1981), *Nature*, 290, 20-26.
79. Derynck R. (1980), *Nature*, 285, 542-546.
80. Gray P.W. (1982), *Nature*, 295, 503-508.
81. Taniguchi T. (1983), *Nature*, 302, 305-310.
82. Bernstein K.E. (1982), *Nature*, 300, 74-76.
83. Bentley D.L. (1980), *Nature*, 288, 730-733.
84. Chang H.C. (1983), *Nature*, 305, 813-815.
85. Das H.K. (1983), *P.N.A.S. USA*, 80, 3543-3547.
86. Schamboeck A. (1983), *N.A.R.*, 11, 8663-8675.
87. Long E.O. (1983), *EMBO J.*, 2, 389-394.
88. Gustafsson K. (1984), *Scand. J. Immunol.*, 19, 91-97.
89. Sodyer R. (1984), *EMBO J.*, 3, 879-885.
90. Strachan T. (1984), *EMBO J.*, 3, 887-894.
91. Law S.W. (1983), *Biochem. Biophys. Res. Commun.*, 112, 257-264.
92. Brewer H.B. (1978), *Biochem. Biophys. Res. Commun.*, 80, 623-630.
93. Karathanasis S.K. (1983), *P.N.A.S. USA*, 80, 6147-6151.
94. Degen S.J.F. (1983), *Biochemistry*, 22, 2087-2097.
95. Chandra T. (1983), *P.N.A.S. USA*, 80, 1845-1848.
96. Rixon M.W. (1983), *Biochemistry*, 22, 3237-3244.
97. Kant J.A. (1983), *P.N.A.S. USA*, 80, 3953-3957.
98. Chung D.W. (1983), *Biochemistry*, 22, 3244-3250.
99. Chung D.W. (1983), *Biochemistry*, 22, 3250-3256.
100. Yang F. (1983), *P.N.A.S. USA*, 80, 5875-5879.
101. Rauegi G. (1983), *N.A.R.*, 11, 5811-5819.
102. Shen L.P. (1982), *P.N.A.S. USA*, 79, 4575-4579.
103. Comb M. (1982), *Nature*, 295, 663-666.
104. Noda M. (1982), *Nature*, 297, 431-434.
105. Horikawa (1983), *Nature*, 306, 611-614.
106. Takahashi H. (1981), *FEBS lett.*, 135, 97-102.
107. Morinaga T. (1983), *P.N.A.S. USA*, 80, 4604-4608.
108. Colantuoni V. (1983), *N.A.R.*, 11, 7769-7776.
109. Wetekam W. (1982), *Gene*, 19, 179-183.
110. Jilka R.L. (1977), *P.N.A.S. USA*, 74, 5692-5696.
111. Jilka R.L. (1979), *J. Biol. Chem.*, 254, 9270-9276.
112. Burstein Y. (1977), *P.N.A.S. USA*, 74, 716-720.
113. Tonegawa S. (1978), *P.N.A.S. USA*, 75, 1485-1489.
114. Burstein Y. (1977), *Biochemistry*, 17, 2392-2400.
115. Smith G.P. (1978), *Biochem. J.*, 171, 337-347.
116. Nishioka Y. (1980), *J. Biol. Chem.*, 255, 369-374.
117. Early P. (1980), *Cell*, 19, 981-992.
118. Litman G.W. (1983), *Nature*, 303, 349-352.

119. Kvist S. (1983), *EMBO J.*, 2, 245-254.
120. Lalanne J.L. (1983), *N.A.R.*, 11, 1567-1577.
121. Hyldig-Nielsen J.J. (1983), *N.A.R.*, 11, 5055-5071.
122. Saito H. (1983), *P.N.A.S. USA*, 80, 5520-5524.
123. Gray A. (1983), *Nature*, 303, 722-725.
124. Yokota T. (1984), *P.N.A.S. USA*, 81, 1070-1074.
125. Uhler M. (1983), *J. Biol. Chem.*, 258, 9444-9453.
126. Notake M. (1983), *FEBS lett.*, 156, 67-71.
127. Shaw G.D. (1983), *N.A.R.*, 11, 555-573.
128. Higashi Y. (1983), *J. Biol. Chem.*, 258, 9522-9529.
129. Gray P.W.C. (1983), *P.N.A.S. USA*, 80, 5842-5846.
130. Fung M.C. (1984), *Nature*, 307, 233-237.
131. Law S.W. (1981), *Nature*, 291, 201-205.
132. Wiebauer K. (1982), *P.N.A.S. USA*, 79, 7077-7078.
133. Panthier J.J. (1982), *Nature*, 298, 90-92.
134. Windass J.D. (1984), *N.A.R.*, 12, 1361-1376.
135. Gaye P. (1977), *Biochem.Biophys.Res.Comm.*, 79, 903-911.
136. Mercier J.C. (1978), *Biochem.Biophys.Res.Comm.*, 82, 1236-1245.
137. Mercier J.C. (1978), *Biochem.Biophys.Res.Comm.*, 85, 662-670.
138. Furutani Y. (1983), *Nature*, 301, 537-540.
139. Davidson J.M. (1982), *Arch.Biochem.Biophys.*, 218, 31-37.
140. Atkinson A. (1983), *Symp. Genetics & Biotechnology*, 198th Meeting of the Genetical Society, London.
141. Kakidani H. (1982), *Nature*, 298, 245-249.
142. Boileau G. (1983), *N.A.R.*, 11, 8063-8071.
143. Yoo O.J. (1982), *P.N.A.S. USA*, 79, 1049-1053.
144. Gaye P. (1982), *Biochimie*, 64, 173-184.
145. Bernstein K.E. (1983), *N.A.R.*, 11, 7205-7214.
146. Atger M. (1979), *Biochem. J.*, 177, 985-988.
147. Haugen D.A. (1977), *Biochem.Biophys.Res.Comm.*, 77, 967-973.
148. MacDonald R.J. (1982), *J. Biol. Chem.*, 257, 9724-9732.
149. Verbanac K.M. (1983), *Arch.Biochem.Biophys.*, 223, 149-157.
150. MacDonald R.J. (1982), *Biochemistry*, 21, 1453-1463.
151. Quinto C. (1982), *P.N.A.S. USA*, 79, 31-35.
152. MacDonald R.J. (1980), *Nature*, 287, 117-122.
153. Villa-Komaroff L. (1978), *P.N.A.S. USA*, 75, 3727-3731.
154. Talmadge K. (1981), *Nature*, 294, 176-178.
155. Sorokin A.V. (1982), *Gene*, 20, 36-46.
156. MacDonald R.J. (1982), *J. Biol. Chem.*, 257, 14582-14585.
157. Seeburg P.H. (1977), *Nature*, 270, 486-494.
158. Ohkubo H. (1983), *P.N.A.S. USA*, 80, 2196-2200.
159. Ivell R. (1984), *P.N.A.S. USA*, 81, 2006-2010.
160. Schmale H. (1983), *EMBO J.*, 2, 763-767.
161. Hudson P. (1981), *Nature*, 291, 127-131.
162. Funckes C.L. (1983), *J. Biol. Chem.*, 258, 8781-8787.
163. Parkes M. (1982), *Nature*, 298, 92-94.
164. Viskochil D.H. (1983), *J. Biol. Chem.*, 258, 8861-8866.
165. Strauss A.W. (1977), *J. Biol. Chem.*, 252, 6846-6855.
166. Sargent T.D. (1981), *P.N.A.S. USA*, 78, 243-246.
167. Deschenes R.J. (1984), *P.N.A.S. USA*, 81, 726-730.
168. McKean D.J. (1978), *Biochemistry*, 17, 5215-5219.
169. Dandekar A.M. (1982), *P.N.A.S. USA*, 79, 3987-3991.
170. Malsky M.L. (1979), *J. Biol. Chem.*, 254, 1580-1585.

171. Chow V. (1983), *FEBS lett.*, 153, 275-279.
172. Ricca G.A. (1981), *J. Biol. Chem.*, 256, 11199-11202.
173. Dijkema R. (1984), *N.A.R.*, 12, 1227-1242.
174. Gordon J.I. (1982), *J. Biol. Chem.*, 257, 971-978.
175. McLean J.W. (1983), *J. Biol. Chem.*, 258, 8993-9000.
176. Kandala J.C. (1983), *N.A.R.*, 11, 3169-3186.
177. Jacobs J.W. (1981), *Science*, 213, 457-459.
178. Chin W.W. (1983), *P.N.A.S.*, 80, 4649-4653.
179. Bar-Nun S. (1980), *P.N.A.S. USA*, 77, 965-969.
180. Botelho L.H. (1979), *J. Biol. Chem.*, 254, 5635-5640.
181. Perler F. (1980), *Cell*, 20, 555-566.
182. Palmiter R.D. (1977), *J. Biol. Chem.*, 252, 6386-6393.
183. Thibodeau S.N. (1978), *J. Biol. Chem.*, 253, 3771-3774.
184. Geiser M. (1983), *J. Biol. Chem.*, 258, 9024-9030.
185. Tate V.E. (1983), *N.A.R.*, 11, 91-104.
186. Chan L. (1980), *J. Biol. Chem.*, 255, 10060-10063.
187. Reynaud G.A. (1983), *P.N.A.S. USA*, 80, 4099-4103.
188. Karr S.R. (1981), *J. Biol. Chem.*, 256, 5946-5949.
189. Sures I. (1984), *P.N.A.S. USA*, 81, 380-384.
190. Hahn V. (1983), *N.A.R.*, 11, 4541-4552.
191. Lund P.K. (1982), *P.N.A.S. USA*, 79, 345-349.
192. Magazin M. (1982), *P.N.A.S. USA*, 79, 5152-5156.
193. Sumikawa K. (1982), *N.A.R.*, 10, 5809-5822.
194. Noda M. (1982), *Nature*, 299, 793-797.
195. Lin Y. (1981), *P.N.A.S. USA*, 78, 2825-2829.
196. Davies P.L. (1982), *P.N.A.S. USA*, 79, 335-339.
197. Suchanek G. (1978), *P.N.A.S. USA*, 75, 701-704.
198. Vlasak R. (1983), *Eur. J. Biochem.*, 135, 123-126.
199. Garfinkel M.D. (1983), *J. Mol. Biol.* 168, 765-789.
200. Hovemann B. (1981), *N.A.R.*, 9, 4721-4734.
201. Crouch M.L. (1983), *J. Mol. Appl. Genet.*, 2, 273-283.
202. Rogers J.C. (1983), *J. Biol. Chem.*, 258, 8169-8174.
203. Slightom J.L. (1983), *P.N.A.S. USA*, 80, 1897-1901.
204. Lycett G.W. (1983), *N.A.R.*, 11, 2367-2380.
205. Higgins T.J.V. (1983), *J. Biol. Chem.*, 258, 9544-9549.
206. Geraghty D. (1981), *N.A.R.*, 9, 5163-5174.
207. Pedersen K. (1982), *Cell*, 29, 1015-1026.
208. Marks D.M. (1982), *J. Biol. Chem.*, 257, 9976-9983.
209. Kurjan J. (1982), *Cell*, 30, 933-943.
210. Singh A. (1983), *N.A.R.*, 11, 4049-4063.
211. Arima K. (1983), *N.A.R.*, 11, 1657-1672.
212. Taussig R. (1983), *N.A.R.*, 11, 1943-1954.
213. Ozaki L.S. (1983), *Cell*, 34, 815-822.
214. Boothroyd J.C. (1981), *N.A.R.*, 9, 4735-4743.
215. Ahmed C.M.I. (1982), *Gene*, 20, 339-346.
216. Watson R.J. (1982), *Science*, 218, 381-384.
217. Eisenberg R.J. (1984), *J. Virol.*, 49, 265-268.
218. Watson R.J. (1983), *Gene*, 26, 307-312.
219. Porter A.G. (1979), *Nature*, 282, 471-477.
220. Min Jou W. (1980), *Cell*, 19, 683-686.
221. Ward C.W. (1979), *Brit. Med. Bull.*, 35, 64-73.
222. Hiti A.L. (1981), *Virology*, 111, 113-124.
223. Verhoeyen M. (1983), *N.A.R.*, 11, 4703-4712.
224. Krystal M. (1982), *P.N.A.S. USA*, 79, 4800-4804.
225. Markoff L. (1982), *Virology*, 119, 288-297.
226. Fields S. (1981), *Nature*, 290, 213-217.

-
227. Hiti A.L. (1982), *J. Virol.*, 41, 730-734.
228. Shaw, M.W. (1982), *P.N.A.S. USA*, 79, 6817-6821.
229. Redmond S.M.S. (1983), *EMBO J.*, 2, 125-131.
230. Yelverton E. (1983), *Science*, 219, 614-620.
231. Schwartz D.E. (1983), *Cell*, 32, 853-859.
232. Both G.W. (1983), *J. Virol.*, 48, 335-339.
233. Lingappa V. (1983), *J. Biol. Chem.*, 253, 8667-8670.
234. Gallione C.J. (1983), *J. Virol.*, 46, 162-169.
235. Kotwal G.J. (1983), *Virology*, 129, 1-11.
236. Skipper N. (1984), *EMBO J.*, 3, 107-111.
237. Bostian K.A. (1984), *Cell*, 36, 741-751.
238. Palva I. (1981), *Gene*, 15, 43-51.
239. Wells J.A. (1983), *N.A.R.*, 11, 7911-7925.
240. Sloma A. (1983), *N.A.R.*, 11, 4997-5004.
241. Mezes P.S.F. (1983), *FEBS lett.*, 161, 195-200.
242. Neugebauer K. (1981), *N.A.R.*, 9, 2577-2588.
243. Stephens M.A. (1984), *J. Bacteriol.*, 158.
244. Yang M. (1983), *N.A.R.*, 11, 237-349.
245. Yamazaki H. (1983), *J. Bacteriol.*, 156, 327-337.
246. Ohmura K. (1983), *Biochem. Biophys. Res. Commun.*, 112, 678-683.
247. McKern N.M. (1983), *FEBS lett.*, 164, 149-153.
248. Kaczorek M. (1983), *Science*, 221, 855-858.
249. Ratti G. (1983), *N.A.R.*, 11, 6589-6595.
250. Yamamoto T. (1982), *J. Bacteriol.*, 152, 506-509.
251. Dallas W.S. (1980), *Nature*, 288, 499-501.
252. Spicer E.K. (1981), *P.N.A.S. USA*, 78, 50-54.
253. So M. (1980), *P.N.A.S. USA*, 77, 4011-4015.
254. Lee C.H. (1983), *Infect. Immun.*, 42, 264-268.
255. Picken R.N. (1983), *Infect. Immun.*, 42, 269-275.
256. Inouye S. (1977), *P.N.A.S. USA*, 74, 1004-1008.
257. Boyd A. (1983), *Nature*, 301, 623-626.
258. Emr S. (1980), *Nature*, 285, 82-85.
259. Hedgepath J. (1980), *P.N.A.S. USA*, 77, 2621-2625.
260. Movva N.R. (1980), *J. Biol. Chem.*, 255, 27-29.
261. Beck E. (1980), *N.A.R.*, 8, 3011-3024.
262. Mizuno T. (1983), *FEBS lett.*, 151, 159-164.
263. Overbeake N. (1983), *J. Mol. Biol.*, 163, 513-532.
264. Michaelis S. (1982), *Annu. Rev. Microbiol.*, 36, 435-465.
265. Hachett J. (1983), *N.A.R.*, 11, 6487-6495.
266. Baga M. (1984), *J. Bacteriol.*, 157, 330-333.
267. Jaurin B. (1981), *Nature*, 290, 221-225.
268. Ambler R.P. (1978), *P.N.A.S. USA*, 75, 3732-3736.
269. Sutcliffe J.G. (1978), *P.N.A.S. USA*, 75, 3737-3741.
270. Postle K. (1983), *P.N.A.S. USA*, 80, 5235-5239.
271. Kikuchi Y. (1981), *N.A.R.*, 9, 5671-5678.
272. Inoye H. (1982), *J. Bacteriol.*, 149, 434-439.
273. Magota K. (1984), *J. Bacteriol.*, 157, 909-917.
274. Bedouille H. (1980), *Nature*, 285, 78-81.
275. Oxender D.L. (1980), *P.N.A.S. USA*, 77, 2005-2009.
276. Wilson V.G. (1980), *J. Biol. Chem.*, 255, 6745-6770.
277. Buchel D.E. (1980), *Nature*, 283, 541-544.
278. Rogers B.L. (1981), *Eur. J. Biochem.*, 116, 165-170.
279. Sugimoto K. (1977), *J. Mol. Biol.*, 111, 487-507.
280. van Wezenbeek P.M.G.F. (1980), *Gene*, 11, 129-148.
281. Wolfe P.B. (1983), *J. Biol. Chem.*, 258, 12073-12080.
-

282. Yamagata (1981), J. Biol. Chem., 256, 2194-2198.
283. Braun G. (1983), Eur. J. Biochem., 137, 495-500.
284. Dunn R. (1981), P.N.A.S. USA, 78, 6744-6748.
285. Vlasuk G.P. (1983), J. Biol. Chem., 258, 7141-7148.
286. Froholm L.O. (1977), FEBS lett., 73, 253-256.
287. Hermodson M.A. (1978), Biochemistry, 17, 442.
288. Sustry P.A. (1983), FEBS lett., 151, 442-445.
289. Minton N. (1983), Symp. Genetics & Biotechnology, 198th Meeting of Genetics Society, London.
290. Freudl R. (1983), Eur. J. Biochem., 134, 497-502.
291. Higgins C.F. (1981), P.N.A.S. USA, 78, 6038-6042.
292. Higgins C.F. (1982), Nature, 298, 723-727.
293. Nakamura K. (1980), P.N.A.S. USA, 77, 1369-1373.
294. Braun C. (1982), N.A.R., 10, 2367-2378.
295. Sako T. (1983), N.A.R., 11, 7679-7693.
296. McLaughlin J.R. (1981), J. Biol. Chem., 256, 11283-11291.
297. Mekalanos J.J. (1983), Nature, 306, 551-557.
298. Lockman H. (1983), J. Biol. Chem., 258, 13722-13726.
299. Viebrock A. (1982), EMBO J., 1, 565-571.
300. Kaput J. (1982), J. Biol. Chem., 257, 15054-15088.
301. Nagata S. (1983), P.N.A.S. USA, 80, 6192-196.
302. Schmidt G.W. (1979), J. Cell. Biol., 83, 615-622.
303. Coruzzi G. (1983), J. Biol. Chem., 258, 1399-1402.
304. Berry-Lowe S.L. (1982), J. Mol. Appl. Genet., 1, 483-498.
305. Smith S.M. (1983), N.A.R., 11, 8719-8734.
306. Broglie R. (1983), Biotechnology, 1, 55-61.
307. Ichihara S. (1982), J. Biol. Chem., 257, 495-500.
308. Benson S.A. (1983), Cell, 32, 1325-1335.
309. Henning U. (1983), Eur. J. Biochem., 136, 233-240.
310. Kadonga J.T. (1984), J. Biol. Chem., 259, 2149-2154.
311. Pollitt S. (1983), J. Bacteriol., 153, 27-32.