9

Sequence analysis of membrane proteins

JOHN E. WALKER and IAN M. FEARNLEY

9.1 Scope of the review

The primary structures of proteins can be determined by two distinct methods: directly by sequence analysis of the protein or indirectly by DNA sequence analysis of the corresponding gene or c-DNA. The direct protein sequence analysis methods are strongly influenced by the nature of the protein. They have been largely developed to accommodate water-soluble globular proteins, and methods for isolation and sequence analysis of hydrophobic proteins are relatively deficient and underdeveloped. The indirect sequence analysis by DNA sequence analysis is not influenced by the nature of the protein.

This advantage of the indirect method notwithstanding, the necessity for the direct analysis of membrane proteins (as for other classes of protein) remains for the following reasons:

(1) Information about post-translational processing of proteins can only be obtained by direct sequence analysis. These processing events include phosphorylation, methylation, glycosylation, acylation and removal of parts of the polypeptide chain by proteolysis.
(2) A prominent method for isolation of genes from libraries in bacteria is based upon the use of synthetic oligonucleotide hybridization probes. These are designed on the basis of protein sequence obtained by direct means.
(3) The interpretation of DNA sequence is aided by an independent knowledge of the sequence of the protein encoded therein. Such knowledge for instance can help to identify start and stop codons and intervening

Techniques for the Analysis of Membrane Proteins. Edited by C. I. Ragan & R. J. Cherry. Published in 1986 by Chapman and Hall, 11 New Fetter Lane, London EC4P 4EE

sequences; it was essential in establishing the existence of overlapping genes (Sanger *et al.*, 1977; Shaw *et al.*, 1978). In analysis of complete genomes (e.g. bacteriophages ϕX174 and lambda, human and bovine mitochondrial DNA) protein sequences were used to identify the genes (Sanger *et al.*, 1977; Shaw *et al.*, 1978; Walker *et al.*, 1980, 1982a).

(4) Indirect and direct protein sequences are obtained independently and so they serve as a useful mutual checking procedure.

Therefore, the first part of this review will describe the extant methods for direct sequencing of membrane proteins including the isolation of the protein, production and fractionation of derived fragments, and their sequence analysis: it will discuss the deficiencies and difficulties of these methods. The second part will outline current methods for isolation of genes for membrane proteins and for the determination of their DNA sequences. It will particularly emphasize the close relationship between the indirect and direct methods and the benefits that accrue from the use of both methods. The review will conclude with a discussion of procedures for prediction of secondary structures of membrane proteins from their amino acid sequences.

9.2 Direct sequence analysis of membrane proteins

9.2.1 WHAT IS A MEMBRANE PROTEIN?

Membrane proteins fall into two broad classes, extrinsic and intrinsic (see Fig. 9.1). The extrinsic proteins are globular structures that lie outside the lipid bilayer and do not penetrate the lipid bilayer at all (e.g. β_2-microglobulin, mitochondrial F_1-ATPase; Type I in Fig. 9.1). A second class of extrinsic membrane proteins (see Table 9.1) are anchored totally or partly via covalent lipid (Type II in Fig. 9.1). The intrinsic membrane proteins are themselves of several kinds. Some have extensive globular domains that lie outside the membrane and are anchored by a hydrophobic membrane sector (Types III, IV

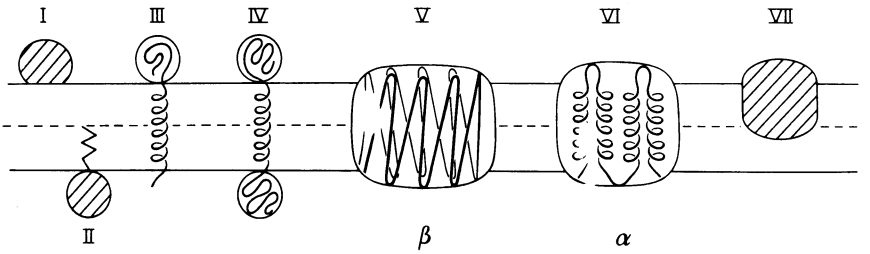

Fig. 9.1 Types of membrane proteins. I. Superficially associated extrinsic membrane protein. II. Extrinsic protein anchored via covalently attached lipid. III and IV. Intrinsic membrane proteins with α-helical membrane span and extensive extramembrane domains. V. Intrinsic membrane protein folded in β-sheets and VI in α-helices with minor extramembrane domains. VII. Partially penetrating membrane protein.

Table 9.1 Membrane-associated proteins with covalent lipid

Protein	*Lipid*	*Footnotes*
E. coli outer membrane lipoprotein	Palmitic acid, *cis*-vaccenic acid, 9,10-methylenehexadecanoic acid, palmitoleic acid	1
c-AMP-dependent protein kinase	α-*N*-Myristoyl	2
Calcineurin B	α-*N*-Myristoyl	3
Murine retrovirus $P15^{gag}$	α-*N*-Myristoyl	4
Rous sarcoma virus p60src	α-*N*-Myristoyl	5–8
Harvey sarcoma virus p21	Palmitoyl	6
Lipophilin	Palmitic acid, stearic acid, oleic acid	9
HLA-B, HLA-DR heavy chains	Palmitic acid	10
Transferrin receptor	Palmitic acid	11
Cytochrome b_5 reductase	α-*N*-Myristoyl	12
Plasmodium falciparum variant surface antigen	*sn*-1,2-Dimyristoylglycerol	13
Trypanosome brucei variant surface antigen	*sn*-1,2-Dimyristoylglycerol	14, 15

Footnotes. 1. Attached via amide to *N*-terminal cysteine and ester linked to glycerol, itself thioether linked to cysteine-1 (Hantke and Braun, 1973). 2. Carr *et al.* (1982). 3. Membrane association has not been demonstrated (Aitken *et al.*, 1982). 4. Henderson *et al.* (1983); Schultz and Oroszlan (1984). 5. Buss and Sefton (1985a,b); Schultz *et al.* (1985). 6. Sefton *et al.* (1982). 7. Garber *et al.* (1983). 8. Pellman *et al.* (1985). 9. Esterified at threonine-198; Stoffel *et al.* (1983). 10. Via thioester bond; Kaufman *et al.* (1984). 11. Omary and Trowbridge (1981). 12. Ozols *et al.* (1984). 13. Haldar *et al.* (1985). 14. Attachment to *C*-terminal oligosaccharide via phosphodiester bond between *sn*-3-glycerolhydroxyl and a sugar hydroxyl; phospholipase *c* releases protein. Ferguson *et al.* (1985). 15. Thy-1, acetylcholinesterase, 5′-nucleotidase and alkaline phosphatase are attached in a similar way. Kolata (1985).

in Fig. 9.1), others are predominantly hydrophobic and are found almost totally within the lipid bilayer (Types V and VI in Fig. 9.1). The extrinsic proteins and the extrinsic protein domains of intrinsic proteins present few difficulties for direct protein sequencing; the major problems are associated with the hydrophobic intrinsic membrane sectors. They are of necessity largely composed of hydrophobic amino acids and are insoluble in aqueous solutions. This discussion is directed towards the primary structural determination of this class of membrane proteins (see Table 9.2).

9.2.2 ISOLATION OF MEMBRANE PROTEINS

(a) *Native proteins*

The methods for the isolation of individual membrane proteins and membrane

Table 9.2 Membrane protein sequences determined by direct protein sequence analysis

Protein	*Source*	*Footnotes*
Glycophorin	Red cell membrane	1
Cytochrome b_5	Liver microsomes	2,3,4
ATP synthase proteolipid	Mitochondria, chloroplasts, bacteria	5
ATP synthase-associated protein	Yeast mitochondria	6
Lipophilin	Human and bovine brain	7–11
Bacteriorhodopsin	*Halobacterium halobium*	12–14
Rhodopsin	Bovine and ovine retina	15,16
Light-harvesting polypeptides	*Rhodospirillaceae*	17–20
Bacteriochlorophyll *c* binding	*Chloroflexus aurantiacus*	21
Cytochrome oxidase subunits	Bovine mitochondria	22–25
Cytochrome c_1	Bovine mitochondria	26
ADP/ATP translocase	Bovine mitochondria	27
Uncoupling protein	Hamster mitochondria	28
H-2K	Mouse lymphocytes	29
Anion transport protein	Human erythrocytes	30
*Omp*F	*E. coli*	31
Transducin γ-subunit	Cow	32
Surface glycoprotein	*Trypanosoma brucei*	33
Asialoglycoprotein receptor	Rat liver	34

Footnotes. 1. Tomita and Marchesi (1975). 2. Takagaki *et al.* (1980). 3. Kondo *et al.* (1979). 4. Ozols and Heinemann (1982). 5. Sebald and Hoppe (1981). 6. Velours *et al.* (1984). 7. Laursen *et al.* (1984). 8. Lees *et al.* (1983). 9. Stoffel *et al.* (1983). 10. Jollès *et al.* (1979). 11. Stoffel *et al.* (1985). 12. Gerber *et al.* (1979). 13. Khorana *et al.* (1979). 14. Ovchinnikov *et al.* (1979). 15. Ovchinnikov (1982). 16. Brett and Findlay (1983). 17. Brunisholz *et al.* (1985). 18. Theiler *et al.* (1984). 19. Brunisholz *et al.* (1984a). 20. Brunisholz *et al.* (1981). 21. Wechsler *et al.* (1985). 22. Tanaka *et al.* (1979). 23. Steffens and Buse (1979). 24. Biewald and Buse (1982). 25. Erdweg and Buse (1985). 26. Wakabayashi *et al.* (1980, 1982). 27. Aquila *et al.* (1982). 28. Aquila *et al.* (1985). 29. Coligan *et al.* (1978). 30. Hydrophobic sector only; Brock *et al.* (1983). 31. Chen *et al.* (1982). 32. McConnell *et al.* (1984). 33. Allen *et al.* (1982). 34. Drickamer *et al.* (1984).

complexes are extremely diverse and a summary of such procedures is beyond the scope of this review (see Nelson and Robinson, 1983). Many of these procedures depend upon the differential extraction of the protein from the membrane under particular conditions of detergent, salt concentration and pH. The choice of detergent is particularly important as is an understanding of the properties of the detergent. For such information see Chapter 3 and reviews by Helenius and Simons (1975), Tanford and Reynolds (1976), Hjelmel and Chrambach (1984) and Zulauf and Rosenbusch (1983).

(b) *Denatured membrane proteins*

The methods for isolation of membrane proteins in their denatured state depend upon two distinctly different approaches to their solubilization. These

are firstly, solubilization in chaotropes (urea, guanidinium hydrochloride) and strong detergents [sodium dodecyl sulphate (SDS)] and secondly, solubilization in organic solvents. The solubilization procedure will often limit and thereby dictate the subsequent purification steps.

(i) *Chaotropes and strong detergents.* Urea is a relatively mild chaotrope and incapable of solubilizing effectively many membrane proteins (e.g. bacteriorhodopsin). However, if solubilization can be achieved, the normal gamut of chromatographic methods that are applicable to globular proteins (gel filtration, ion-exchange chromatography, reverse phase chromatography) can be employed. Guanidinium hydrochloride is a more effective chaotrope but its use precludes many purification methods, and the choice is restricted to size fractionation. However, this may be effective for the resolution of simple mixtures. The use of sodium dodecyl sulphate also excludes some methods (ion exchange) but it has been used to effect separation of subunits of mitochondrial complexes such as cytochrome oxidase (Steffens and Buse, 1976) and ATP synthase (Steffens *et al.*, 1982). In combination with polyacrylamide gel electrophoresis and isoelectric focusing it provides an extremely powerful method with high resolution for the production of small quantities of highly purified proteins that can be successfully sequenced in modern instruments. The rest of this section will assess this approach.

The method (see Fig. 9.2) consists of the application of the sample dissolved

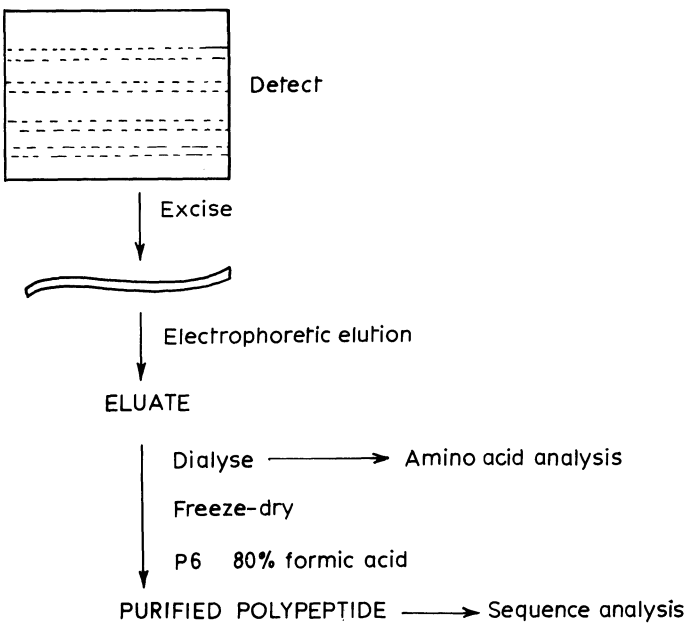

Fig. 9.2 Preparative polyacrylamide gel electrophoresis.

in the presence of SDS to a polyacrylamide slab separating gel surmounted by a stacking gel. The constituent proteins are separated under standard conditions, detected, excised and eluted by electrophoresis into a dialysis bag or similar device. Following dialysis and removal of low-molecular-weight contaminants, the recovered protein can then be sequenced (Bhown *et al.*, 1980; Walker *et al.*, 1982b; Hunkapiller *et al.*, 1983a).

The main considerations in this type of procedure are to minimize chemical modification of the protein during the procedures, to avoid cleavage of acid-sensitive peptide bonds during staining procedures to detect the protein, to maximize recovery and to remove contaminants that would interfere with the subsequent sequencing reactions.

Chemical modifications can be avoided in part by using highly purified reagents for electrophoresis. The introduction of reducing agents (0.1 mM sodium thioglycolate) into the upper chamber (Hunkapiller *et al.*, 1983a) helps to scavenge amino-reactive contaminants, oxidants and free radicals that remain in the gel after polymerization. Excess persulphate can be removed by pre-electrophoresis but only at the cost of the loss of stacking gel and its attendant advantages.

Many methods have been tried for visualization of proteins. These include tagging the protein with fluorescent reagents (fluorescamine) or with radioactivity, by intrinsic labelling or by ^{125}I-iodination. These methods may not be generally applicable (intrinsic labelling) or may alter the mobility of the partially reacted protein relative to the unreacted protein (fluorescamine; Talbot and Yphantis, 1971). Other mild staining procedures employ sodium acetate (Higgins and Dahmus, 1979) or potassium chloride (Nelles and Bamburg, 1976) but are relatively insensitive (Walker *et al.*, 1980). The staining of side bands with Coomassie Blue dye can also be employed, but swelling or shrinkage of the stained side band relative to the unstained gel and irregularities in unstained bands can make precise localization of unstained bands difficult. The method of choice is probably brief staining with Coomassie Blue dye (5 min) followed by destaining for 10–15 min. This allows rapid and precise localization of protein bands which can then be excised with a razor.

The protein dye complex can then be recovered from the gel by passive elution in buffer or, better, by electroelution. Maceration of gel prior to these procedures leads to release of acrylamide oligomers from the gel. Therefore, it should be avoided and the integrity of the gel slices maintained.

Following elution, contaminants should be removed prior to analysis of protein samples. Dialysis against 10% ethanol helps to remove low-molecular-weight components and SDS. However, for small amounts of protein it may lead to further loss on the dialysis membrane surface.

An alternative procedure is to remove the contaminants from the freeze-dried electroeluent by gel filtration in 80% formic acid on a small column of Biogel P6 (Walker *et al.*, 1980). Alternatively, an ethanol precipitation step

(90% ethanol at $-20°$ C; M. Hunkapiller, personal communication) or a chloroform step (Wessel and Flügge, 1984) have been recommended to purify proteins prior to sequence analysis on a gas phase sequencer.

Another procedure used in conjunction with solid phase sequencing (see below) selectively immobilizes the protein by covalent attachment to glass derivatives leaving the low-molecular-weight contaminants in solution. They can then be removed by a simple washing step (Walker *et al.*, 1982b).

Recent developments in this methodology have concentrated on the direct elution of the protein from the gel on to a solid matrix (such as the glass-fibre disc employed as a support in the gas phase sequencer; Aebersold *et al.*, 1986). Such methods avoid a number of steps, are therefore simpler and lead to an overall improvement in recovery and, in addition, allow even smaller amounts of protein to be manipulated. Such methods clearly hold much promise and will be developed further.

(ii) *Gel permeation chromatography in organic solvents.* In 1951 Folch and Lees reported that a group of proteins could be extracted from brain tissue into mixtures of chloroform and methanol. In this property these proteins resemble lipids and hence the proteins were called proteolipids. Later, Cattell *et al.* (1970, 1971) demonstrated that proteolipids could be extracted from bovine mitochondria in a similar manner. Fractionation of this brain extract by gel permeation chromatography on hydroxypropylated Sephadex G-25 (Sephadex LH-20) was first reported by Mokrasch (1967). The proteins themselves were not resolved from each other and were found in the excluded volume of the column. However, further fractionation of the proteins was achieved by partition chromatography (see below) on the same support. Sandermann and Strominger (1972) showed that hydroxypropylated Sephadex G-50 (Sephadex LH-60) had a more effective separation range for proteins in their purification of a lipophilic enzyme C_{55}-isoprenoid alcohol phosphokinase. Later Blondin (1979a) showed that the chloroform/methanol-soluble extract of bovine mitochondria could be separated into three fractions by size fractionation on Sephadex LH-60. These fractions themselves are still very complex (I. M. Fearnley and J. E. Walker, unpublished work).

This simple gel filtration technique has proved to be very useful in the purification of a wide range of proteins extracted from membranes with chloroform and methanol. These include a proteolipid associated with kidney ($Na^+ + K^+$)-ATPase (Reeves *et al.*, 1980) and components of chromatophores of the *Rhodospirillaceae* (Tonn *et al.*, 1977; Theiler *et al.*, 1984; Brunisholz *et al.*, 1981, 1984a,b 1985). The range of the technique has also been extended by Theiler *et al.* (1983) who separated the L and M subunits of photosynthetic reaction centres of *Rhodospirillum rubrum* by gel filtration in organic solvents on hydroxypropylated Sephadex G-100 prepared according to the procedure of Ellingboe *et al.* (1969). Other solvents have also been used for protein

purification in combination with lipophilic Sephadex. For example phospholamban and another proteolipid were isolated from sarcoplasmic reticulum proteolipids by gel filtration on Sephadex LH-60 in 88% formic acid/95% ethanol (1:3, v/v) (Collins *et al.*, 1981).

(iii) *Ion exchange chromatography in organic solvents.* Fractionation of chloroform/methanol-soluble proteolipids has also been achieved by ion-exchange chromatography in the same solvent to which buffer (ammonium acetate) has been added. The procedure is derived from a method devised for fractionation of lipids (Rouser *et al.*, 1969). It is likely that the separation achieved depends not only on ionic interactions but also on partition chromatography. It is important that the ion-exchange matrices are pretreated as recommended by Rouser *et al.* (1969). (See also Fillingame, 1976, and Theiler *et al.*, 1984). The method has been used as a purification step in the isolation of the carbodiimide-reactive proteolipid (Fillingame, 1976; Graf and Sebald, 1978; Sebald *et al.*, 1979) and a second proteolipid (Velours *et al.*, 1982, 1984) from ATP synthases, and in the purification of light-harvesting proteins from *Rhodospirillaceae* (Theiler *et al.*, 1983, 1984).

(iv) *Normal phase and reverse phase chromatography.* Lipophilic Sephadex swells only in polar organic solvents, so polar components of a solvent mixture become more strongly associated with it thereby creating a marked difference in composition between stationary and mobile phases. Therefore, separations can be achieved because the more polar components of a mixture of solutes are more strongly associated with the stationary phase. LH-20 and LH-60 Sephadex are suitable materials for partition (normal phase) chromatography. They have been used in this way to fractionate proteolipids from brain (Soto *et al.*, 1969) and from beef mitochondria (Cattell *et al.*, 1970, 1971). Earlier, brain proteolipids were also fractionated by partition chromatography on silicic acid (Matsumoto *et al.*, 1964). Rather surprisingly, this kind of chromatography has been little used for membrane protein fractionation under the highly controlled conditions of high-performance liquid chromatography. Rubinstein (1979) has emphasized the usefulness of the approach for chromatography of hydrophobic proteins. He demonstrated the use of Lichrosorb Diol as a suitable support in the fractionation of hydrophobic protein components of foetal calf serum and also demonstrated a wider applicability of the procedure in the isolation of human leucocyte interferon (Rubinstein *et al.*, 1978).

Reverse phase chromatography, which employs hydrophobic matrices to preferentially bind the less polar components of solvent mixtures, has been used to great effect for purification of globular proteins and subfragments. However, the strong and often irreversible binding of hydrophobic membrane proteins to this kind of support has severely restricted its usefulness as a

technique for membrane protein fractionation. Nonetheless, a small number of chloroform/methanol-soluble proteolipids have been purified on C_{18} reverse phase supports. Examples are small proteolipids associated with yeast (Velours *et al.*, 1982, 1984) and bovine ATP synthases (Blondin, 1979a,b; I. M. Fearnley and J. E. Walker, unpublished work) and proteolipids from cardiac membrane (Capony *et al.*, 1983). Tarr and Crabb (1983) found that cytochrome *P*-450 and rhodopsin could be recovered from a CN-bonded phase with ternary solvents containing acetonitrile, propan-1-ol and 0.1% trifluoroacetic acid.

9.2.3 FRAGMENTATION OF MEMBRANE PROTEINS

(a) *Proteolytic enzymes*

Proteolytic enzymes can be employed in two distinctly different ways for degradation of membrane proteins. (1) By treating native membrane proteins (either in the membrane itself or in a dispersed form in mild detergent) limited proteolysis will occur such that the extrinsic membrane sectors are degraded and removed, leaving intact the membrane sectors which are protected by the lipid bilayer. (2) Extensive proteolysis of the entire protein (including the intrinsic membrane sector) can be achieved in the presence of strong detergents (e.g. sodium dodecyl sulphate) or chaotropes (urea, guanidinium hydrochloride).

(i) *Limited proteolysis.* Treatment of glycophorin with trypsin leads to the formation of an insoluble, hydrophobic precipitate. This proved to be the membrane-spanning segment of 32 amino acids in the *C*-terminal half of the molecule. It was shown to anchor the protein in the erythrocyte membrane (Segrest *et al.*, 1972; Tomita and Marchesi, 1975). Since then a growing number of membrane proteins anchored by a single membrane-spanning segment have been identified by similar means. In a number of cases these segments have been isolated and their sequences determined by direct sequence analysis. In numerous examples where the sequences of the proteins have been determined by DNA sequencing (see Table 9.3) the positions and sequences of membrane anchors have been proposed by looking for hydrophobic spans within the predicted protein sequences (see Section 9.4.1). Examples are influenza virus neuraminidase (Fields *et al.*, 1981) the *b* subunit of ATP synthase (Gay and Walker, 1981) and surface glycoproteins of Epstein–Barr virus (Baer *et al.*, 1984; Hudson *et al.*, 1985).

Proteolytic cleavage of proteins in the membrane proved to be a helpful procedure in the elucidation of the structure of rhodopsin. A number of studies had shown that a range of proteases cleaved the native protein into two large membrane-bound fragments. (For example see Pober and Stryer, 1975; Sale *et*

al., 1977; Towner *et al.*, 1977; Hargrave *et al.*, 1980; Mullen and Akhtar, 1981). Brett and Findlay (1983) and Ovchinnikov *et al.* (1983) in their sequence analysis of ovine and bovine rhodopsins put this finding to good effect. They cleaved the native rhodopsin with proteases to produce two membrane-bound fragments of M_r 27 K and 12 K respectively. These were separated by chromatography, further fragmented and sequenced. A similar approach was also valuable in the structural analysis of bacteriorhodopsin (Gerber *et al.*, 1979; Ovchinnikov *et al.*, 1979).

Limited proteolysis of membrane proteins is not only valuable in the generation of large fragments for sequence analysis, but also gives valuable information about the topography of the protein molecule: degraded segments are accessible and lie outside the membrane, protected (undegraded) segments lie within the lipid bilayer. In the case of bacteriorhodopsin, proteolysis experiments showed that the *C*-terminal 23 amino acids, seven residues at the *N*-terminal end of the chain and a short segment between residues 70 and 77 were accessible to proteolysis and therefore exposed to the aqueous milieu (Gerber *et al.*, 1979; Ovchinnikov *et al.*, 1979; Walker *et al.*, 1979).

(ii) *Proteolysis in denaturing conditions.* Many of the specific endoproteinases commonly used for digestion of globular proteases are active in the presence of SDS (0.1%) or urea (2–8 M); so they can be employed for degradation of membrane proteins in the presence of denaturants. A demonstration of the stability of many endoproteases in the presence of SDS is proteolytic digestion of proteins in polyacrylamide gels (Cleveland *et al.*, 1977). The list of proteinases that can be used in this way include trypsin, chymotrypsin, staphylococcal V8 proteinase, subtilisin, thermolysin, endoproteinase lysC and clostripain. Of the exoproteinases carboxypeptidase Y is most stable, being active in the presence of 0.1% SDS. So providing susceptible bonds are present, and the proteins can be solubilized, the extensive proteolysis of membrane proteins is not a serious problem; the difficulty is the purification of the peptides that are generated (see Section 9.2.4). Other procedures have also been recommended (Allen, 1981). For example, succinylation, maleylation or citraconylation of amino groups in membrane proteins produces polyanionic derivatives that often are soluble at alkaline pH in the absence of denaturants and so can be proteolysed easily. Another procedure that has found occasional use for digestion of membrane proteins is to first solubilize in concentrated (90–100%) formic acid and then to digest with pepsin (Allen, 1980a; Hayes *et al.*, 1986).

(b) *Chemical cleavage*

Membrane proteins are usually soluble in formic and trifluoroacetic acids (70–100%) and so cleavage on the *C*-terminal side of methionyl bonds with cyanogen bromide has been widely used for their fragmentation. (For

examples see Ovchinnikov *et al.*, 1979; Gerber *et al.*, 1979; Brett and Findlay, 1983.) Cleavage on the *C*-terminal side of tryptophanyl bonds has also proved useful. A number of different reagents have been employed for this purpose. For example, Green and Toms (1985) cleaved Ca^{2+}-ATPase according to the procedure of Savige and Fontana (1977). The protein was first carboxymethylated, delipidated in the presence of SDS and then the freeze-dried protein–SDS complex dispersed by sonication in the cleavage reagent (containing acetic acid, 12 M HCl, dimethyl sulphoxide and phenol). Brunisholz *et al.* (1985) used iodosobenzoic acid to cleave light-harvesting proteins from *Rps. viridis* and Lees *et al.* (1983) employed 2-(2-nitrophenylsulphenyl)-3-methyl-3′-bromoindolenine (NBS-skatole) to produce fragments of a brain proteolipid. Additional cleavages after tyrosine as well as tryptophan residues can be made with the more powerful reagent, *N*-bromosuccinimide (Khorana *et al.*, 1979).

Cleavage of asparagine–glycine bonds with hydroxylamine can be performed in the presence of 6 M-guanidinium hydrochloride and partial hydrolysis with acid (predominant cleavage of aspartyl–X bands) could also in principle be used for cleavage of membrane proteins.

9.2.4 PURIFICATION OF FRAGMENTS DERIVED FROM MEMBRANE PROTEINS

The purification of fragments from digests of membrane proteins has been a major obstacle in their sequence analysis. The difficulties are to solubilize and to disaggregate the fragments under conditions that do not chemically modify the peptides and which permit high-resolution chromatography. The range of chromatographic methods available is in essence that described in Section 9.2.2 for the purification of membrane proteins. However, the problems of disaggregation and solubilization associated with the complex mixtures of peptides of rather similar chemical characteristics that result from fragmentations have often proved to be more severe than those encountered in protein purification. So greater effort has been invested in this problem, and a wider range of chromatographic parameters have been investigated. These methods are described in the following sections.

(a) *Gel permeation chromatography*

Size fractionations have been performed successfully on mixtures of hydrophobic peptides with combinations of Sephadex and 6 M-guanidinium hydrochloride (Ovchinnikov *et al.*, 1979) and polyamide supports (Biogels) with various dilutions with water or acetic acid (Erdweg and Buse, 1985; Brunisholz *et al.*, 1985; Lees *et al.*, 1983) or formic acid (Steffens and Buse, 1976; Jolles *et al.*, 1979). Lipophilic Sephadex (LH-20 or LH-60) has been employed with mixtures of formic acid and ethanol (Gerber *et al.*, 1979; Takagaki *et al.*, 1980; Green and Toms, 1985) or with a mixture of formic acid, acetic acid, chloroform and ethanol (Brett and Findlay, 1983). Allen (1977)

used solvents containing phenol, acetic acid and water in attempts to disaggregate and resolve on Sephadex mixtures of hydrophobic peptides from Ca^{2+}-ATPase, and phenol/water/formamide mixtures for subsequent chromatographic steps. Unfortunately the peptides recovered at the end of these purification steps had blocked *N*-terminals, having been modified during chromatography.

(b) *High-performance liquid chromatography*

Formic acid, in water and in combination with various other solvents (usually alcohols), has been used as a solvent for purification of hydrophobic peptides by high-performance chromatography. It was introduced by Gerber *et al.* (1979) and Takagaki *et al.* (1980) for isolation of fragments of bacteriorhodopsin and cytochrome b_5 by reverse phase chromatography on C_{18} supports. Subsequently, Velours *et al.* (1984) used formic acid and propanol with a C_{18} matrix in their sequence analysis of an ATPase-associated proteolipid from yeast. Stoffel *et al.* (1982a,b, 1985) in their work on lipophilins have preferred silica supports with similar solvents. Ternary solvents containing formic acid, trimethylamine and propanol with cyanopropyl supports have been advocated by Tarr and Crabb (1983) for isolation of large hydrophobic peptides.

The major drawback of all these methods is the inimical nature of formic acid to protein, instruments, columns and experimenters. Tarr and Crabb (1983) and Walker *et al.* (1985) have noted that exposure of proteins and peptides to this acid causes chemical modification and thus leads to additional chemical complexity. For example, serine and threonine residues become esterified, a reaction that may be reversed by treatment with aminoethanol (Tarr and Crabb, 1983).

A somewhat different approach was used by Aquila *et al.* (1982). They maleylated the ADP/ATP carrier of bovine mitochondria and citraconylated a fragment of it, and then performed partition chromatography of blocked fragments on silica. Peptides not purified by this method were rechromatographed on sulphopropyl-Sephadex in pyridine acetate gradients containing propan-1-ol.

It has been possible to isolate some hydrophobic peptides under conditions employed for more hydrophilic fragments. For example the four cyanogen bromide fragments of the A6L proteolipid from bovine mitochondria can be purified by reverse phase chromatography on a C_{18} support in trifluoracetic acid and acetonitrile (I. M. Fearnley and J. E. Walker, unpublished work).

(c) *Gas chromatography of polyaminoalcohols*

Volatility of peptides can be increased by methylation, trifluoroacetylation and reduction with $LiAl[^2H]_4$ or $B_2[^2H]_6$ to the polyaminoalcohols (Kelley *et al.*, 1975; Nau and Biemann, 1976; Frank *et al.*, 1978). The resultant derivatives can then be purified by gas chromatography which can be linked directly to a

mass spectrometer (see Section 9.2.5.c). This method has been used in the sequence analysis of bacteriorhodopsin (Gerber *et al.*, 1979) and lipophilin (Stoffel *et al.*, 1982a).

9.2.5 SEQUENCE ANALYSIS

Most protein sequences determined directly on the protein (see Table 9.2) have depended largely upon the Edman degradation whereby amino acids are removed sequentially from the *N*-terminal of the protein or polypeptide (for reviews see Allen, 1981; Edman and Henschen, 1975). This is achieved by a cycle of reactions containing two steps (see Fig. 9.3):

(1) *coupling* of the *N*-terminal residue with phenylisothiocyanate (PITC);
(2) *cleavage* of the *N*-terminal residue by a cyclization reaction in acid;

In a third reaction, *conversion*, the cleavage product, the anilinothiazolinone (ATZ), is rearranged to a more stable form, the phenylthiohydantoin (PTH). This derivative is then identified, usually by high-performance liquid chromatography. A manual version of the procedure has been widely used in various modified forms (see Allen, 1981) particularly for degradation of small peptides. However, these procedures have been largely superseded by automated procedures performed in specially designed instruments, protein (or peptide) sequencers. They are of three types: the liquid phase (or spinning cup) sequencer, the solid phase sequencer and the gas phase sequencer.

(a) *The spinning cup sequencer*

This instrument was designed by Edman and Begg (1967). The heart of the instrument is a glass beaker (or cup) rotated about its vertical axis in a controlled environment. The protein is introduced in solution and forms a film on the wall of the cup. It is dried *in vacuo*, redissolved in a non-volatile buffer, Quadrol, and reacted in this film with PITC. Excess reagents, buffer and products are then removed by extraction of the film with a solvent (a mixture of benzene and ethyl acetate). The protein film is again dried. Cleavage is then effected with an anhydrous acid and then the ATZ removed by extraction with butyl chloride. This extract is collected, dried, converted to the PTH in a second reaction chamber (the convertor) and then collected for subsequent identification. Meanwhile, the film in the cup, now containing the protein with one amino acid removed from its *N*-terminal continues into the next cycle of degradation.

From the point of view of sequence analysis of membrane proteins the major disadvantage of this instrument is that the hydrophobic membrane proteins have a propensity for dissolving in the organic solvents (ethyl acetate, benzene and butyl chloride) and so these proteins tend to be extracted from the cup and lost. This problem can be alleviated by making the protein more polar, for

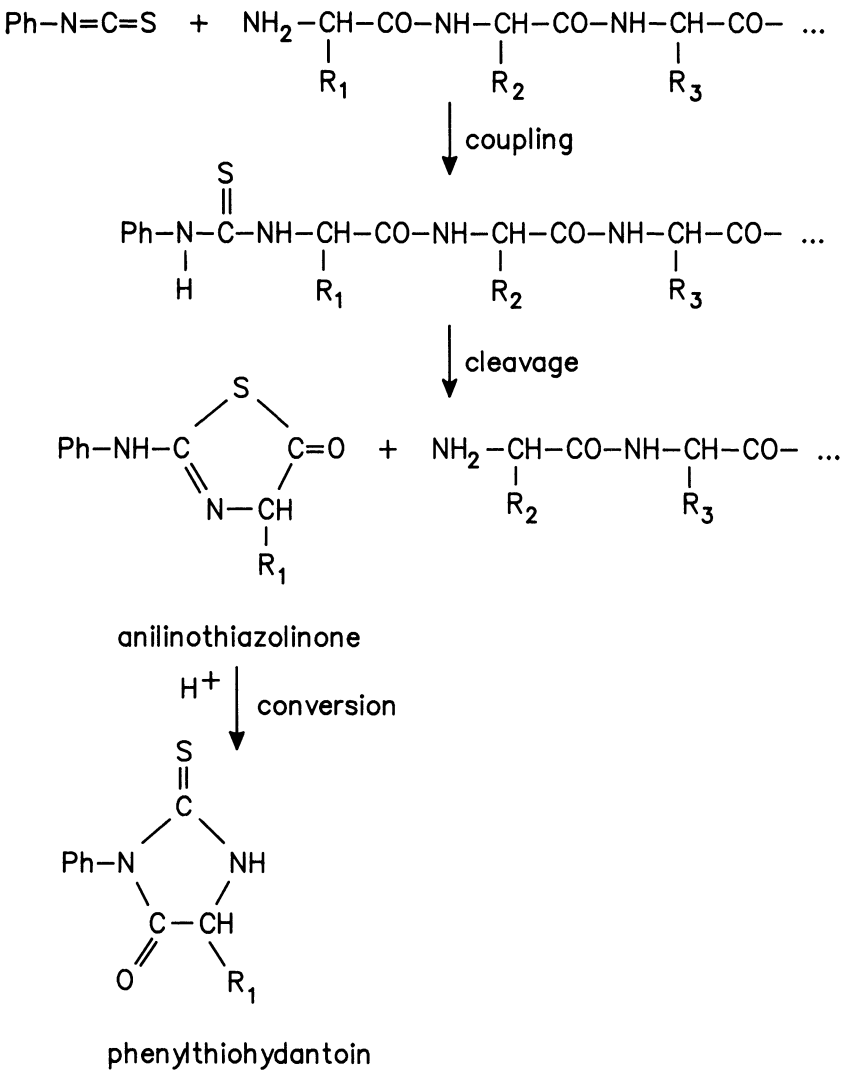

Fig. 9.3 The Edman degradation.

example by reaction in the first cycle of the Edman degradation with polar sulphonated isothiocyanates that will thereby derivatize ε-amino groups with polar residues. The α-NH_2 group also reacts but is removed by the cleavage reaction (Braunitzer *et al.*, 1970, 1971). Other hydrophilic derivatives can also be made by carbodiimide coupling of carboxyl groups with hydrophilic amines (Foster *et al.*, 1973) or by reaction of the *C*-terminal homoserine lactone of CNBr peptides with polar derivatives (Braunitzer and Pfletschinger, 1978). However, the charged amino acids tend to be rare in membrane proteins and

so coupling of polar groups to carboxyls and ε-amino groups have been little used amongst this group of proteins. However, the reaction of the *C*-terminal lactone of CNBr peptides with solid supports (rather than polar derivatives) has been widely used for immobilization of cyanogen bromide peptides for solid phase sequence analysis (see Section 9.2.5.b).

An important device for reducing the wash out of proteins and peptides from the film of the spinning cup sequencer is the addition of the polycationic compound, Polybrene (Tarr *et al.*, 1978; Klapper *et al.*, 1978). Its use has extended greatly the range of peptides and proteins that can be degraded in the spinning cup sequencer, and it is equally important in the gas phase instrument (Section 9.2.5.c).

These potential difficulties notwithstanding the liquid phase sequencer has been used successfully for sequencing a number of membrane proteins (e.g. Brunisholz *et al.*, 1984a,b, 1985). These experiments probably owe their success to the fact that although these proteins contain long hydrophobic stretches, they do also have polar amino acids in their sequences and particularly arginine residues near to their *C*-terminals. This, and the presence of Polybrene, apparently are sufficient to retain the protein in the reaction cup. However, hydrophobic proteins lacking such polar features tend not to be retained in the reaction cup and other means have to be employed to determine their sequences.

(b) *The solid phase sequencer*

An alternative solution to the problem of sample loss in extraction solvents incurred in the spinning cup sequencer, is the covalent attachment of the protein or peptide to an insoluble support. This is the basis of the solid phase sequencer (Laursen, 1971). The support with the attached peptide is placed in a small column (the counterpart of the cup in the spinning cup sequencer). Reagents can then be introduced sequentially as required by the Edman degradation. The major advantage is that by-products can be removed by solvent washes, without concomitant danger of loss of the sample.

A range of supports has been proposed (see Laursen *et al.*, 1975). However, those based on porous glass are most valuable and have found widest use. They have excellent physical properties and can be easily derivatized to give covalently bound amino groups (see Fig. 9.4). These in turn can be activated to isothiocyanato derivatives which will react with amino groups of a protein or can be acylated with the *C*-terminal homoserine lactone (Horn and Laursen, 1973) or spirolactone (Wachter and Werhahn, 1979) generated by chemical cleavage on the *C*-terminal sides of methionyl and tryptophanyl groups, respectively. Attachment via carboxyl groups activated with water-soluble carbodiimides has proved to be more problematical, but has found some use.

This approach to sequence analysis of membrane proteins has been particularly successful and has been employed extensively in the sequence

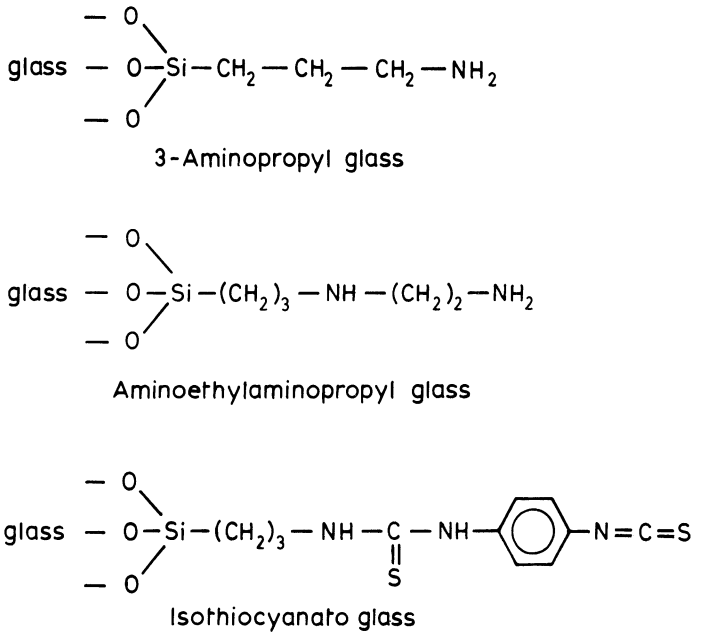

Fig. 9.4 Glass derivatives for protein immobilization. See Wachter *et al.* (1975).

analysis of the most hydrophobic proteins such as bacteriorhodopsin (Gerber *et al.*, 1979), the membrane segment of cytochrome b_5 (Takagaki *et al.*, 1980) and the proteolipids associated with ATP synthases from various sources (Hoppe and Sebald, 1980; Sebald and Hoppe, 1981; Velours *et al.*, 1984; I. M. Fearnley and J. E. Walker, unpublished results).

(c) *The gas phase sequencer*

The gas phase sequencer (Hunkapiller and Hood, 1980; Hewick *et al.*, 1981, Hunkapiller *et al.*, 1983b) differs from the liquid phase and solid phase sequencers in two important aspects. Firstly, two reagents, the buffer (trimethylamine) for the coupling reaction and the acid for cleavage (trifluoroacetic acid) are delivered as vapours in an argon stream. Thereby large excesses of reagents are avoided and side products and contaminants are minimized. As in the other designs of sequencer the PITC is delivered as a liquid, and washes with appropriate solvents (ethyl acetate, heptane, butyl chloride) are interposed between coupling and cleavage and after cleavage to remove reaction by-products and ATZ respectively. Secondly, the reaction chamber in this instrument is a cartridge in which a glass fibre disc is sandwiched. The disc is impregnated with Polybrene and the samples applied to this matrix. This instrument has been particularly successful for micro-sequence analysis in the range of 10 pmol–1 nmol. Hitherto it has been little used for analysis of hydrophobic proteins and direct comparisons of

sequencing efficiency of the same hydrophobic proteins in gas and solid phase instruments have not been made. Successful attempts have been made to sequence peptides covalently attached to glass beads in this instrument, as a way of avoiding sample loss (Strickler *et al.*, 1984).

(d) *C-terminal sequences*
In the case of membrane proteins with exposed *C*-terminal domains, *C*-terminal sequence can be deduced by treatment with carboxypeptidases A and B (Ambler, 1972), C (Tschesche and Kupfer, 1972) or Y (Martin *et al.*, 1977; see Allen, 1981). An example of this is the determination of the sequence of the *C*-terminal region of bacteriorhodopsin (Gerber *et al.*, 1979; Ovchinnikov *et al.*, 1979). Denatured membrane proteins can be investigated in the presence of SDS with carboxypeptidase Y.

(e) *Mass spectrometry*
Electron impact mass spectrometry has been used to deduce sequences of peptides and in rare cases of complete proteins (see Morris and Dell, 1975). The major problem has been to convert the peptides to volatile derivatives; this has been achieved by acetylation and permethylation (Morris *et al.*, 1971), reduction to polyaminoalcohols (Nau, 1976; Frank and Desiderio, 1978) and trimethylsilylation. In combination with gas chromatographic separation of polyaminoalcohols the method has been successfully used in the sequence analyses of the membrane proteins bacteriorhodopsin and lipophilin (see Section 9.2.4.c). The more recently developed technique of fast atom bombardment mass spectrometry (Barber *et al.*, 1981) was employed to identify myristic acid as the hydrophobic membrane anchor of a number of proteins that become associated with membrane (Aitken *et al.*, 1982; Carr *et al.*, 1982). Gas chromatography/electron impact mass spectrometry was employed to identify the same blocking groups in retrovirus proteins (Henderson *et al.*, 1983). Amongst membrane proteins, as with other proteins, these techniques have much promise and utility, particularly when used in combination with other sequencing methods (Gibson and Biemann, 1984).

9.3 Indirect methods of protein sequencing

An ever increasing number of protein sequences are being determined by DNA sequence analysis of genes and complementary DNA. The major difficulty in this approach is the isolation of the desired clones containing the DNA sequence encoding the protein of interest. (For detailed reviews on these techniques see Maniatis *et al.*, 1982.) The methods employed to do this depend strongly upon the origin of the protein (gene) as summarized in Fig. 9.5. DNA from organelles (mitochondria, chloroplasts) and some viruses is often not very large and so the entire genome can be cloned and sequenced. In the case

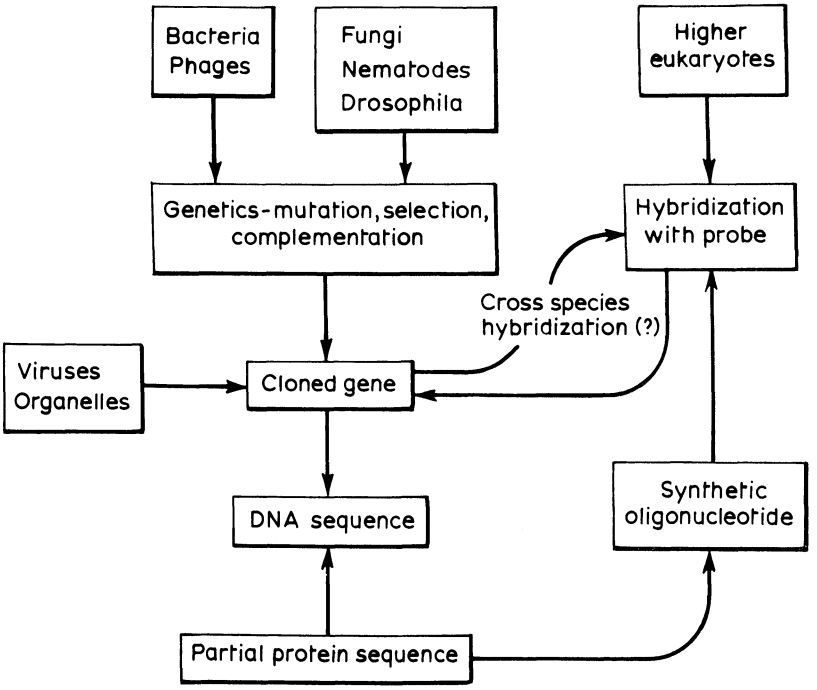

Fig. 9.5 Strategies for gene cloning.

of genes from prokaryotic organisms and bacteriophages, viruses and simpler eukaryotes (e.g. fungi, nematodes, *Drosophila*), the powerful tools of genetics can be employed to isolate genes. However, these tools cannot be applied to the isolation of nuclear genes from higher eukaryotes and so different techniques are used. These can also be equally well applied to prokaryotic DNA. Three general methods are available.

(1) Expression methods in which the expression of a gene *in vivo* or *in vitro* is detected by biological assay or by immunochemical techniques. (For a discussion of these methods see Maniatis *et al.*, 1982.)
(2) Mixed oligonucleotide probes, 11–17 nucleotides in length, are synthesized to correspond to (fragmentary) protein sequence information. The degeneracy of the genetic code is accounted for by incorporation of a mixture of two to four bases (if necessary) at particular positions. These oligonucleotides are labelled with [^{32}P]phosphate and used as hybridization probes under stringent hybridization conditions (Wood *et al.*, 1985) to identify, in libraries, individual recombinants containing the sequences of interest (reviewed by Smith, 1983).
(3) An alternative procedure employs long (50–90 bases) unique oligonucleotide hybridization probes (Anderson and Kingston, 1983). In this case the

Table 9.3 Membrane protein sequences determined by DNA sequencing

Function	*Protein*	*Source*	*Comments*	*Footnotes*
1. Energy transduction				
(a) ATP synthase components	Subunit *a*	E. coli	Component of proton channel	1
	ATPase-6	*Synechococcus* 6301	Equivalent to bacterial *a*	2
		D. melanogaster (mt)		3,4
		Mammals (mt)		5–7
		A. nidulans (mt)		8
		S. cerevisiae (mt)		9
	Subunit X	Pea (ch)	Equivalent to bacterial *a*	10
	Subunit *b*	*E. coli*	Component of H^+-channel, binds F_1	1
		Synechococcus 6301	Also has a duplicated version, b′.	2
	Subunit I	Wheat (ch)	Equivalent to bacterial *b*	11
	Subunit *c*	*E. coli*	Component of proton channel	1
		Synechococcus 6301		2
	ATPase-9	Cow (n)	Equivalent to bacterial *c*	12
		A. nidulans (mt)		13
		N. crassa (n)		14
		S. cerevisiae (mt)		15
	Subunit III	Spinach (ch)	Equivalent to bacterial *c*	16
		Wheat (ch)		17
	aap I	*S. cerevisiae* (mt)	Associated with ATP synthase	18
		A. nidulans (mt)		19
	A6L	*D. melanogaster* (mt)	Associated with ATP synthase	4
		Mammals (mt)		5–7
(b) Cytochrome bc_1	Cytochrome *b*	*A. nidulans* (mt)		19
		Mammals (mt)		5–7
		S. cerevisiae (mt)		20
		Z. mays (mt)		21
	Cytochrome c_1	*S. cerevisiae* (n)		22

Table 9.3 (continued)

Function	*Protein*	*Source*	*Comments*	*Footnotes*
(c) Cytochrome oxidase	Subunit I	*A. nidulans* (mt)		23
		D. melanogaster (mt)		4
		Mammals (mt)		5–7
		S. cerevisiae (mt)		24
	Subunit II	*D. melanogaster* (mt)		4
		Mammals (mt)		5–7
		Z. mays (mt)		25
		S. cerevisiae (mt)		26,27
	Subunit III	*D. melanogaster* (mt)		4
		Mammals (mt)		5–7
		S. cerevisiae (mt)		28
(d) Oxidoreductases				
NADH dehydrogenase		Mammals (Mt)	6 subunits of complex	5–7,29
		E. coli	One chain only	30
Succinate dehydrogenase	4 subunits	*E. coli*	sdhA, *sdh*B, *sdh*C, *sdh*B	31,32
Fumarate reductase	4 subunits	*E. coli*	*frd*A, *frd*B, *frd*C, *frd*D	33,34
(e) Photosynthesis	Reaction centre subunits L,H,M	*Rps. sphaeroides*		35
	light-harvesting subunits α,β	*Rps. capsulata*		35a
	Chlorophyll a	Pea (n)	Chlorophyll binding	36,37
	Apoprotein	Spinach (ch)		38
	P680 chlorophyll a apoprotein	Spinach (ch)		39

Table 9.3 (continued)

Function	*Protein*	*Source*	*Comments*	*Footnotes*
	Ps II 32 kD protein	Tobacco (ch)	Psb or herbicide-binding protein of	40
		Spinach (ch)	PsII plastoquinone reductase	40
	PsII 39 kD protein	Spinach (ch)	homologous to PsbA	38
(e) Photosynthesis	Cytochrome *f*	pea (ch)	Component of b_6f complex	41
		Spinach (ch)		42
		Wheat (ch)		43
	15.2 kD protein	Pea (ch)	Component of b_6f complex	44
		Spinach (ch)		45
	Apocytochrome b_6	Spinach (ch)	Component of b_6f complex	45
	Bacteriorhodopsin	*H. halobium*	Light-driven proton pump	46
2. Immune response				
	IgM			47
MHC antigen complex	class I	Human, mouse		48
	class II	Human, mouse		49
Cell surface antigens	MRC OX-2 glycoprotein	Mouse		50
	Thy-1	Rat, mouse		51
3. Receptors				
(a) Acetylcholine	$\alpha,\beta,\gamma,\delta$	Calf muscle		52–56
	α,γ	Human		52,57
	α,β,γ	*Torpedo californica*		58–61
	α	*Torpedo marmorata*		62,63
(b) Asialoglycoprotein	Asialoglycoprotein receptor		Glycoprotein transport	64,65
(c) Epidermal growth factor	EGF-receptor	Human		66
(d) Insulin	Insulin receptor	Human		67
(e) Interleukin 2	Interleukin 2 receptor	Human		68

Table 9.3 (continued)

Function	*Protein*	*Source*	*Comments*	*Footnotes*
(f) LDL	Low-density lipoprotein receptor	Human	Cholesterol transport	69
(g) T-cell	α,β	Human		70–72
(h) Transferrin	Transferrin receptor	Human Murine		73 74
(i) IgR	Immunoglobulin receptor	Rabbit	Transcellular transporter of IgA and IgM	75
4. Transport				
(a) Cations	Ca^{2+}-ATPase	Rabbit sarcoplasmic reticulum	Ca^{2+} and Mg^{2+} vectorial ion transport	76
	$(Na^+ + K^+)$-ATPase	Sheep kidney		77
	α subunit	*Torpedo californica*	Na^+ and K^+ vectorial exchange	78
	Kdp-ATPase	*E. coli*	ATP-driven K^+ transport	79
	Na^+ channel	*Electrophorus*	Na^+ permeability of electrically excitable membranes	80
(b) Anions	Band III protein	Mouse	Cl^-/HCO_3^- exchange in erythrocyte membrane	81
	*Pst*A, *Pst*B	*E. coli* inner membrane	Phosphate-specific transport	82
(c) Amino acids	*His*P, *His*Q, *His*M	*E. coli* inner membrane	High-affinity histidine transport	83
(d) Sugars	Lactose permease	*E. coli* inner membrane	*lac*Y	84
	Glucose transporter	Human erythrocyte		85
	Maltose transporter	*E. coli* inner membrane	*mal*F	86

Table 9.3 (continued)

Function	*Protein*	*Source*	*Comments*	*Footnotes*
4. Transport				
(e) Nucleotides	ADP/ATP translocase	*N. crassa* mitochondria (n)	Translocation of ADP and ATP in and out of mitochondria	87
		Z. mays mitochondria (n)		88
(f) Pores	*Omp*C	*E. coli* outer membrane		89
	*Omp*F	*E. coli* outer membrane	Porin	90
	*Lam*B	*E. coli* outer membrane	Permeation of maltose and maltodextrins, used as lambda λ receptor	91
	*pho*E	*E. coli* outer membrane		92
	Mitochondrial porin	*S. cerevisiae* outer mt-membrane (n)		93
5. Viral	Haemagglutinin	Human influenza		94–96
		Fowl plague influenza		97
	Neuraminidase	Human influenza		98,99
	Glycoprotein	Vesicular stomatitis		100
	Glycoproteins	Epstein–Barr		101,102
	Variable surface glycoprotein	*Trypanosoma brucei*		103
	E1 glycoprotein	Corona virus		104
	E1, E2 glycoproteins	Semliki Forest virus		105
6. Miscellaneous				
(a) Cytochrome *P*-450	P450MC	Rat		106
	$P_1$450	Mouse, human		107,108
	$P_3$450	Mouse		107
	P450 (SSC)	Cow		109
(b) Surface antigens		*Plasmodium falciparum*		110
(c) Visual proteins	Transducin α-subunit	Bovine		111
	Rhodopsin	Human		112

Table 9.3 (continued)

Function	*Protein*	*Source*	*Comments*	*Footnotes*
(d) Myelin sheath protein	Po glycoprotein	Rat Schwann cell		113
(e) Flagellar rotation	*motA*	*E. coli*		114
7. Unknown	Glycoproteins	Epstein–Barr virus		101
	Interferon-induced P16	Human		115
	uncI	*E. coli*		1

Footnotes. 1. Gay and Walker (1981); 2. A. L. Cozens and J. E. Walker, unpublished work; 3. mt, mitochondrial gene; n, nuclear gene; ch, chloroplast gene; 4. de Bruijn (1983); 5. Anderson *et al.* (1981); 6. Anderson *et al.* (1982); 7. Bibb *et al.* (1981); 8. Grisi *et al.* (1982); 9. Macino and Tzagoloff (1980); 10. Cozens *et al.* (1986); 11. Bird *et al.* (1985); 12. Gay and Walker (1985); 13. Brown *et al.* (1984); 14. Viebrok *et al.* (1982); 15. Macino and Tzagoloff (1979); 16. Alt *et al.* (1983); 17. Howe *et al.* (1982); 18. Macreadie *et al.* (1983); 19. Waring *et al.* (1981); 20. Nobrega and Tzagoloff (1980); 21. Dawson *et al.* (1984); 22. Sadler *et al.* (1984); 23. Waring *et al.* (1984); 24. Bonitz *et al.* (1980); 25. Fox and Leaver (1981); 26. Fox (1979); 27. Coruzzi and Tzagoloff (1979); 28. Thalenfield and Tzagoloff (1980); 29. Chomyn *et al.* (1985); 30. Young *et al.* (1981); 31. Wood *et al.* (1984); 32. Darlison and Guest (1984); 33. Grundström and Jaurin (1982); 34. Cole *et al.* (1982); 35. Sutton *et al.* (1982); 35a. Youvan *et al.* (1984); 36. Corruzzi *et al.* (1983); 37. Cashmore (1984); 38. Alt *et al.* (1984); 39. Morris and Herrmann (1984); 40. Zurawski *et al.* (1982); 41. Willey *et al.* (1984a). 42. Alt and Herrmann (1984); 43. Willey *et al.* (1984b); 44. Phillips and Gray (1984); 45. Heinemeyer *et al.* (1984); 46. Dunn *et al.* (1981); 47. Rogers *et al.* (1980); 48. Ploegh *et al.* (1981); 49. Kaufman *et al.* (1984); 50. Clark *et al.* (1985); 51. Seki *et al.* (1985); 52. Noda *et al.* (1983c); 53. Tanabe *et al.* (1984); 54. Takei *et al.* (1984); 55. Takei *et al.* (1985); 56. Kubo *et al.* (1985); 57. Shibahara *et al.* (1985); 58. Noda *et al.* (1982); 59. Noda *et al.* (1983a); 60. Noda *et al.* (1983b); 61. Claudio *et al.* (1983); 62. Devillers-Thiery *et al.* (1983); 63. Sumikawa *et al.* (1982); 64. Spiess and Lodish (1985); 65. Spiess *et al.* (1985); 66. Ullrich *et al.* (1984); 67. Ullrich *et al.* (1985); 68. Cosman *et al.* (1984); 69. Sudhof *et al.* (1985); 70. Sim *et al.* (1984); 71. Yanagi *et al.* (1984); 72. Hedrick *et al.* (1984); 73. Schneider *et al.* (1984); 74. Stearne *et al.* (1985); 75. Mostov *et al.* (1984); 76. MacLennan *et al.* (1985); 77. Shull *et al.* (1985); 78. Kawakami *et al.* (1985); 79. Hesse *et al.* (1984); 80. Noda *et al.* (1984). 81. Kopito and Lodish (1985); 82. Surin *et al.* (1984); 83. Higgins *et al.* (1982); 84. Büchel *et al.* (1980); 85. Mueckler *et al.* (1985); 86. Froshauer and Beckwith (1984); 87. Arrends and Seband (1984); 88. Baker and Leaver (1985); 89. Mizuno *et al.* (1983); 90. Inokuchi *et al.* (1982); 91. Clement and Hofnung (1981); 92. Overbeeke *et al.* (1983); 93. Mihara and Sato (1985); 94. Winter *et al.* (1981); 95. Verhoeyen *et al.* (1980); 96. Gething *et al.* (1980); 97. Porter *et al.* (1979); 98. Blok and Air (1982); 99. Fields *et al.* (1981); 100. Rose *et al.* (1980); 101. Baer *et al.* (1984); 102. Hudson *et al.* (1985); 103. Boothroyd *et al.* (1982); 104. Armstrong *et al.* (1984); 105. Garoff *et al.* (1980); 106. Yabusaki *et al.* (1984); 107. Kimura *et al.* (1984); 108. Jaiswal *et al.* (1985); 109. Morohashi *et al.* (1984); 110. Holder *et al.* (1985); 111. Tanabe *et al.* (1985); 112. Nathans and Hogness (1984); 113. Lemke and Axel (1985); 114. Dean *et al.* (1984); 115. T. Evans and D. S. Secher, unpublished.

degeneracy of the genetic code is accounted for by 'guessing' the most likely unique codon for a particular amino acid. The guesses are influenced by a knowledge of codon usage (codon strategy) in the organism from which the DNA originates, of the relative stability of G-T versus A-C mismatches and of the infrequency of the dinucleotide dCpG in eukaryotic genes (Caruthers, 1985).

(4) A recent development is the use of synthetic probes containing deoxyinosine at ambiguous codon positions. It is proposed that this analogue is an inert base that neither destabilizes nor contributes at mismatched sites towards formation of a DNA duplex (Takahashi *et al.*, 1985).

These methods, and particularly methods 2–4, require an input of protein sequence independently determined by direct means.

Having obtained a clone of interest, the determination of the recombinant DNA sequence by the dideoxy method coupled with cloning in bacteriophage M13 (Sanger *et al.*, 1977; Biggin *et al.*, 1983; Bankier and Barrell, 1983) or by the chemical method (Maxam and Gilbert, 1977) is extremely rapid.

Thus, a modern, efficient approach to sequence analysis of proteins (irrespective of their nature) combines partial direct protein sequence information with the indirectly determined information deduced from DNA sequences. This approach has revolutionized the sequence analysis of membrane proteins (as other proteins) and the majority of the sequences of membrane proteins now known have been determined by this approach (see Table 9.3).

9.4 Secondary structures of membrane proteins

9.4.1 SECONDARY STRUCTURE PREDICTION

Membrane proteins contact two distinct environments, the lipid phase in which they are embedded, and the aqueous phase into which they protrude. It is to be expected that the parts of the protein in contact with the lipid will be made of hydrophobic amino acids and that those protruding into the aqueous environment will be more polar. However, some polar amino acids may be functional (for example in an ion channel) or structural, and so may also be within the lipid bilayer. In the latter case they will probably pair with opposite charges to form salt bridges. It is also to be expected that the parts of membrane proteins buried in the hydrophobic interior of the membrane will be largely folded in α-helices or β-sheets. This is because these arrangements allow the hydrophilic CO and NH groups of the polypeptide backbone to form hydrogen bonds. In the absence of such bonds to each other they would need to form

hydrogen bonds with water. By the same argument, less regular parts of the chain containing bends required to link α-helices or β-sheets should be either at the hydrophilic membrane surface or in the aqueous environment, where the groups can form hydrogen bonds (Unwin and Henderson, 1984). Two kinds of membrane proteins have been observed so far; bacteriorhodopsin and photosynthetic reaction centres are made largely of transmembrane α-helices, and bacterial porin is made of β-sheets.

These simple considerations and the physical thickness of the lipid bilayer (ca. 40 Å) impose constraints on membrane protein structure that can be exploited in the prediction of their secondary structures from primary sequences. This is particularly true for transmembrane α-helices. They will be made of stretches of predominantly hydrophobic acids. With a vertical rise in an α-helix of 1.5 Å/amino acid residue these stretches will be 25–27 amino acids in length. It can also be anticipated that these α-helical segments will be joined by more hydrophilic structures containing turns (β-turns).

Various methods of calculating hydrophobicity along a sequence have been proposed (Segrest and Feldmann, 1974; Kyte and Doolittle, 1982; Engelman *et al.*, 1982). The most widely used today employs a computer program SOAP to calculate hydrophobicity (Kyte and Doolittle, 1982; for a FORTRAN listing see Rose *et al.*, 1985). For this purpose a hydropathy scale, an amalgam of experimental observations has been composed that takes into consideration the hydrophilic and hydrophobic properties of each amino acid. The program continuously determines the average hydropathy within a segment of predetermined length as it advances through the sequence, and the consecutive scores are plotted from *N*- to *C*-terminus. As an example, the program finds seven hydrophobic segments in the sequence of bacteriorhodopsin (Fig. 9.6). These are assumed to correspond to α-helical membrane spans.

Numerous other examples of the use of this program to predict the positions of transmembrane segments in amino acid sequences are to be found in the literature. Some predictions can be developed further by correlation with other biochemical and structural features (see the ADP/ATP translocase, Saraste and Walker, 1982; cytochromes *b* and b_{563}/f, Saraste, 1984; Widger *et al.*, 1984; Mansfield and Anderson, 1985; Ca^{2+}-ATPase, MacLennan *et al.*, 1985; bacteriorhodopsin, Engleman *et al.*, 1980).

This approach to looking for membrane-spanning segments does not work in the case of the *E. coli* outer membrane protein, porin, as the protein does not contain sizeable hydrophobic domains and consists predominantly of β-structures. About two-thirds of the polypeptide backbone is arranged in antiparallel β-pleated sheet in an orientation approximately normal to the membrane plane, the average strand length being 10–12 residues (Kleffel *et al.*, 1985). However, the extents of these segments can be predicted by looking in the primary structure for sequences that promote turns and thereby reverse the direction of the polypeptide chain (Paul and Rosenbusch, 1985).

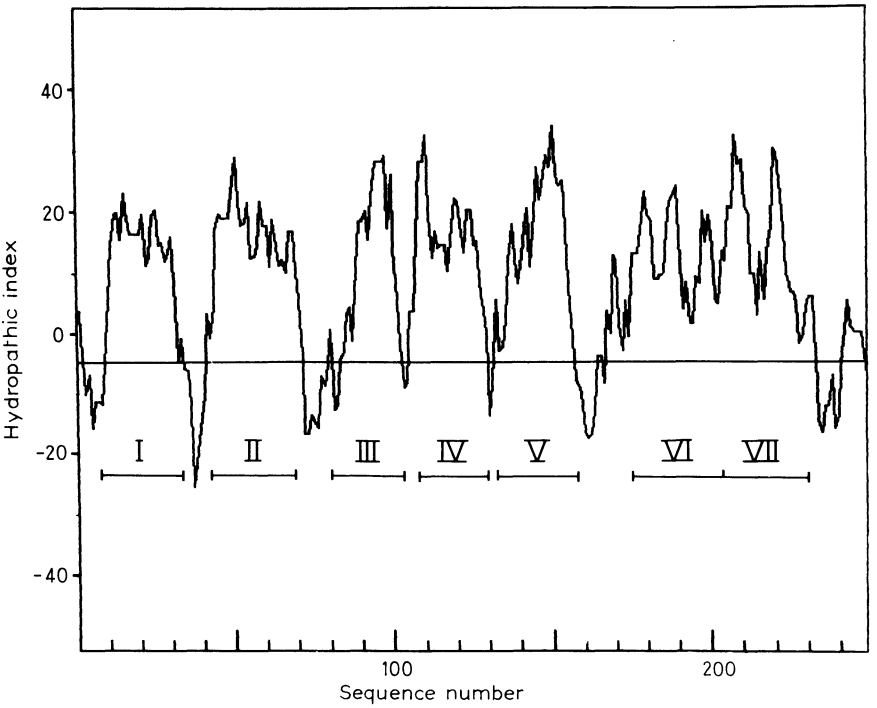

Fig. 9.6 Hydrophobic profile of bacteriorhodopsin. The calculation was made with a span of seven amino acids using the program SOAP (Kyte and Doolittle, 1982). The bars indicate the seven hydrophobic stretches which are assumed to correspond to membrane-spanning segments.

Application of this approach to bacteriorhodopsin produces a series of segments joined by β-turns. Mostly, the segments correspond quite well to hydrophobic segments predicted by the Kyte and Doolittle method (Paul and Rosenbusch, 1985). So this procedure can be viewed as a useful complement to the hydrophobicity calculations. The prediction of β-turns depends upon a knowledge of the amino acids observed in β-turns in known structures, an approach used also for α-helices and β-sheets by Chou and Fasman (1978). It should be noted that the Chou and Fasman method depends upon an analysis of structures of *globular* proteins and so makes falacious predictions of secondary structure when applied to membrane proteins.

The problem of predicting from the proposed helices of bacteriorhodopsin the most likely packing of the helices in the membrane has been discussed by Engleman *et al.* (1980, 1982). An important consideration in their analysis was that buried charges should be satisfied by proximal charges with opposite sign.

In order to gain insight into the possible formation of the ion channel in the acetylcholine receptor, amphipathic features of the predicted membrane-spanning segments were detected by Fourier analysis of the hydrophobicities

of the proteins (Finer-Moore and Stroud, 1984). Amphipathic features are also evident in helical wheel plots (Schiffer and Edmundson, 1967).

9.4.2 CIS α-HELICES

So far in this discussion it has been assumed that protein domains that penetrate the lipid bilayer will span it. The reason for this assumption is that membrane proteins, in which the polypeptide chain penetrates only part way into the membrane and then exits on the same side of the membrane, necessarily require turns or bends to occur in the centre of membrane. This would expose hydrogen-bonding groups in the turn to lipid, an energetically unfavourable arrangement (Henderson, 1981).

Models of lipophilin deduced from primary sequences (Laursen *et al.*, 1984; Stoffel *et al.*, 1984) have been proposed which contain such *cis*-membrane domains. It is argued that this arrangement is partly justified by the presence of turns in the hydrophobic interiors of some globular proteins (Rose *et al.*, 1983). However, these particular turns have bound buried water to satisfy the hydrogen bonds in the buried turns. It has also been proposed that the membrane anchor of cytochrome b_5 has a *cis* arrangement (Takagaki *et al.*, 1983). However, the available data can also be reconciled with a different structure in which the membrane anchor forms an amphipathic helix which lies in the surface of the membrane, its hydrophobic face towards the lipid. Melittin has been shown to bind to membranes in this way (Eisenberg, 1984).

The major value of this predictive approach to secondary structural analysis of membrane proteins is that it proposes experimentally testable models. It also provides the basis for the development of more refined models such as those for bacteriorhodopsin (Engleman *et al.*, 1980, 1982) or cytochrome *b* (Saraste, 1984) which incorporate biochemical as well as primary sequence information. The most generally relevant biochemical information is of two kinds: firstly that which defines the exposed regions of the membrane protein, and secondly, that which defines the regions of the polypeptide that are within the lipid bilayer. The use of proteolytic enzymes for defining exposed regions has already been described (Section 9.2.3.a.i); the following section describes chemical labelling studies of topography of membrane proteins.

9.4.3 CHEMICAL STUDIES OF TOPOGRAPHY

Surface labelling was introduced to study the red cell membrane. In such experiments, the intention is to chemically label only the exposed parts of membrane proteins on the side of the membrane to which the reagent is applied. It is based upon the use of chemical reagents that can label proteins but whose properties prevent them from diffusing through the lipid bilayer. This procedure was originally used by Maddy (1964) who employed a

fluorescent agent, stilbene-4-acetamido-4′-thiocyanate disulphonate (SITS) to label red blood cells. Then Berg (1969) developed ^{35}S-labelled diazobenzene sulphonate, and Bretscher (1971) used [^{35}S]formylmethionylsulphone methyl phosphate to investigate the erythrocyte membrane; pyridoxal phosphate was also used for the same purpose (Rifkin *et al.*, 1972). Lactoperoxidase-generated I^+ has also been employed, but doubt has been cast on its usefulness as extensive labelling of lipids has been reported (Bretscher, 1973).

A more recent and important development has been the introduction of photoactivatable hydrophobic or amphipathic reagents for the general labelling of the hydrophobic core of membranes (for reviews see Brunner, 1981; Bayley, 1982, 1983). The hydrophobic reagents dissolve and react with the entire lipid phase whereas the amphipathic probes react with the membrane such that only the non-polar portion of the molecule penetrates the lipid bilayer. Thus, labelling is restricted to a certain depth in the membrane. The aim is to identify polypeptide segments of integral membrane proteins that are buried within the lipid bilayer.

An example of a reagent that is finding extensive use is 3-trifluoromethyl-3-(*m*-[^{125}I]iodophenyl)diazirine, [^{125}I]TID (Brunner *et al.*, 1979; Brunner and Semenza, 1981; Frielle *et al.*, 1982; Spiess *et al.*, 1982; Jorgensen and Brunner, 1983; Stieger *et al.*, 1984; Hoppe *et al.*, 1984; Kahan and Moscarello, 1985). The basic photoreactive unit, 3-trifluoromethyl-3-phenyldiazirine, has also been incorporated into a fatty acyl chain of a phospholipid which can be used to label from a single leaflet of the lipid bilayer.

The main relevance of this technique to this review is that once the protein has been labelled, the reacted side chains of amino acids must be identified. Once again this requires isolation, fragmentation and direct sequence analysis of the protein. The labelling information may then be used to refine secondary structure predictions.

Further information on topographical studies of membrane proteins is contained in Chapters 6 and 7 of this volume.

9.5 Conclusions and perspectives

An important objective of this review is to demonstrate that the difficulties of direct sequence analysis of membrane proteins can be largely circumvented by the use of a combination of direct protein sequencing and analysis of appropriate DNA fragments. Nonetheless, as emphasized both at the outset and in the preceding section some kinds of information can only be obtained by direct study of the protein. However, the enormous recent growth of primary structural information of membrane proteins is not being paralleled by an increase in secondary and tertiary structural data. Methods for predicting secondary structures of membrane proteins can provide useful information, but lack the precision and conviction of true experimental methods. Significant

advances have been made recently in membrane protein structure, particularly in the development of methods for crystallization of membrane proteins. This has led to the elucidation of a high-resolution model for bacterial photosynthetic reaction centres (Deisenhofer *et al.*, 1984). It can be confidently expected that other structures will follow in the next few years and that they will greatly extend our understanding of membrane biochemistry.

Acknowledgement

We thank R. Henderson for discussions and Miss M. Whittaker for her help in preparation of the manuscript.

9.6 References

Aebersold, R. H., Teplow, D. B., Hood, L. E. and Kent, S. B. H. (1986) *J. Biol. Chem.*, **261**, 4229–38.

Aitken, A., Cohen, P., Santikarn, S., Williams, D. H., Calder, A. G., Smith, A. and Klee, C. B. (1982) *FEBS Lett.*, **150**, 314–18.

Allen, G. (1977) *FEBS 11th Symp.*, **45**, 159–68.

Allen, G. (1980) *Biochem. J.*, **187**, 545–63.

Allen, G. (1981) *Sequencing of Proteins and Peptides* (eds T. S. Work and R. H. Burdon), North-Holland, Amsterdam, New York and Oxford, pp. 266–7.

Allen, G., Gurnett, L. P. and Cross, G. A. M. (1982) *J. Mol. Biol.*, **157**, 527–46.

Alt, J. and Herrmann, R. G. (1984) *Curr. Genet.*, **8**, 551–7.

Alt, J., Morris, J., Werlhoff, P. and Herrmann, R. G. (1984) *Curr. Genet.*, **8**, 597–606.

Alt, J., Winter, P., Sebald, W., Moser, J. G., Schedel, R., Westhoff, P. and Herrmann, R. G. (1983) *Curr. Genet.*, **7**, 129–33.

Ambler, R. (1972) *Methods Enzymol.*, **25**, 143–54.

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. and Young, I. G. (1981) *Nature (London)*, **290**, 457–65.

Anderson, S., de Bruijn, M. H. L., Coulson, A. R., Eperon, I. C., Sanger, F. and Young, I. G. (1982) *J. Mol. Biol.*, **156**, 683–717.

Anderson, S. and Kingston, I. B. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 6838–42.

Aquila, H., Link, T. A. and Klingenberg, M. (1985) *EMBO J.*, **4**, 2369–76.

Aquila, H., Misra, D., Eulitz, M. and Klingenberg, M. (1982) *Hoppe Seyler's Z. Physiol. Chem.*, **363**, 345–9.

Armstrong, J., Nieman, H., Smeekens, S., Rottier, P. and Warren, G. (1984) *Nature (London)*, **308**, 751–2.

Arrends, H. and Sebald, W. (1984) *EMBO J.*, **3**, 377–82.

Baer, R., Bankier, A. T., Biggin, M. D., Deininger, P. L., Farrell, P. J., Gibson, T. J., Hatfull, G., Hudson, G. S., Satchwell, S. C., Seguin, C., Tuffnell, P. S. and Barrell, B. G. (1984) *Nature (London)*, **310**, 207–11.

Baker, A. and Leaver, C. J. (1985) *Nucleic Acids Res.*, **13**, 5857–67.

Bankier, A. and Barrell, B. G. (1983) In *Techniques in Nucleic Acid Biochemistry* (ed R. A. Flavell), Elsevier Scientific Publishers, Amsterdam, pp. B508/1-B508/31.

Barber, M., Bordoli, R. S., Sedgwick, R. D. and Tyler, A. N. (1981) *Nature (London)*, **293**, 270–5.

Bayley, H. (1982) *Membr. Transp.*, **1**, 185–94.

Bayley, H. (1983) *Photogenerated Reagents in Biochemistry and Molecular Biology*, Elsevier Biomedical Press, Amsterdam, pp. 1–187.

Berg, H. C. (1969) *Biochim. Biophys. Acta*, **183**, 65–78.

Bhown, A. S., Mole, J. E., Hunter, F. and Bennett, C. J. (1980) *Anal. Biochem.*, **103**, 184–90.

Bibb, M. J., van Etten, R. A., Wright, C. T., Walberg, M. W. and Clayton, D. A. (1981) *Cell*, **26**, 167–80.

Biewald, R. and Buse, G. (1982) *Hoppe-Seylers Z. Physiol. Chem.*, **363**, 1141–53.

Biggin, M. D., Gibson, T. J. and Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963–5.

Bird, C. R., Koller, B., Auffret, A. D., Huttly, A. K., Howe, C. J., Dyer, T. A. and Gray, J. C. (1985) *EMBO J.*, **4**, 1381–8.

Blok, J. and Air, G. M. (1982) *Biochemistry*, **21**, 4001–7.

Blondin, G. A. (1979a) *Biochem. Biophys. Res. Commun.*, **87**, 1087–94.

Blondin, G. A. (1979b) *Biochem. Biophys. Res. Commun.*, **90**, 355–61.

Bonitz, S. G., Corruzzi, G., Thalenfield, B. E., Tzagoloff, A. and Macino, G. (1980) *J. Biol. Chem.*, **255**, 11927–41.

Boothroyd, J. C., Paynter, A., Coleman, S. L. and Cross, G. A. M. (1982) *J. Mol. Biol.*, **157**, 547–56.

Braunitzer, G. and Pfletschinger, J. (1978) *Hoppe Seyler's Z. Physiol. Chem.*, **359**, 1015–18.

Braunitzer, G., Schrank, B. and Ruhfus, A. (1970) *Hoppe Seyler's Z. Physiol. Chem.*, **351**, 1589–90.

Braunitzer, G., Schrank, B., Ruhfus, A., Petersen, S. and Petersen, U. (1971) *Hoppe Seyler's Z. Physiol. Chem.*, **352**, 1730–2.

Bretscher, M. S. (1971) *J. Mol. Biol.*, **58**, 775–81.

Bretscher, M. S. (1973) *Science*, **181**, 622–9.

Brett, M. and Findlay, J. B. C. (1983) *Biochem. J.*, **211**, 661–70.

Brock, C. J., Tanner, M. J. A. and Kempf, C. (1983) *Biochem. J.*, **213**, 577–86.

Brown, T. A., Ray, J. A., Waring, R. B., Scazzocchio, C. and Davies, R. W. (1984) *Curr Genet.*, **8**, 489–92.

Brunner, J. (1981) *Trends Biochem. Sci.*, **6**, 44–6.

Brunner, J., Hauser, H., Braun, G., Wilson, K. J., Wacker, H., O'Neill, B. and Semenza, G. (1979) *J. Biol. Chem.*, **254**, 1821–8.

Brunner, J. and Semenza, G. (1981) *Biochemistry*, **20**, 7174–82.

Brunisholz, R. A., Cuendet, P. A., Theiler, R. and Zuber, H. (1981) *FEBS Lett.*, **129**, 150–4.

Brunisholz, R. A., Jay, F., Suter, F. and Zuber, H. (1985) *Biol. Chem. Hoppe-Seyler*, **366**, 87–98.

Brunisholz, R. A., Suter, F. and Zuber, H. (1984a) *Hoppe-Seyler's Z. Physiol. Chem.*, **365**, 675–88.

Brunisholz, R. A., Wiemken, V., Suter, F., Bachofen, R. and Zuber, H. (1984b) *Hoppe-Seyler's Z. Physiol. Chem.*, **365**, 689–701.

Büchel, D. E., Gronenborn, B. and Müller-Hill, B. (1980) *Nature (London)*, **283**, 541–5.

Buss, J. E. and Sefton, B. M. (1985a) *J. Virol.*, **53**, 7–12.
Buss, J. E. and Sefton, B. M. (1985b) *Biochemistry*, **20**, 7174–82.
Capony, J. P., Rinaldi, M. L., Guilleux, F. and Demaille, J. G. (1983) *Biochim. Biophys. Acta*, **728**, 83–91.
Carr, S. A., Biemann, K., Shoji, S., Parmelee, D. C. and Titani, K. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6128–31.
Caruthers, M. H. (1985) *Science*, **230**, 281–5.
Cashmore, A. R. (1984) *Proc. Natl. Acad. Sci.*, **81**, 2960–4.
Cattell, K. J., Knight, I. G., Lindop, C. R. and Beechey, R. B. (1970) *Biochem. J.*, **117**, 1011–13.
Cattell, K. J., Lindop, C. R., Knight, I. G. and Beechey, R. B. (1971) *Biochem. J.*, **125**, 169–77.
Chen, R., Krämer, C., Schmidmayr, W., Chen-Schmeisser, U. and Henning, U. (1982) *Biochem. J.*, **203**, 33–43.
Chomyn, A., Mariottini, P., Cleeter, M. W. J., Ragan, C. I., Matsuno-Yagi, A., Hatefi, Y., Doolittle, R. F. and Attardi, G. (1985) *Nature (London)*, **314**, 529–97.
Chou, P. Y. and Fasman, G. D. (1978) *Annu. Rev. Biochem.*, **47**, 251–76.
Clark, M. J., Gagnon, J., Williams, A. F. and Barclay, A. N. (1985) *EMBO J.*, **4**, 113–18.
Claudio, T., Ballivet, M., Patrick, J. and Heinemann, S. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 1111–15.
Clement, J. M. and Hofnung, M. (1981) *Cell*, **27**, 507–18.
Cleveland, D. W., Fischer, S. G., Kirschener, M. W. and Laemmli, U. K. (1977) *J. Biol. Chem.*, **252**, 1102–6.
Cole, S. T., Grundström, T., Jaurin, B., Robinson, J. J. and Weiner, J. H. (1982) *Eur. J. Biochem.*, **126**, 211–16.
Coligan, J. E., Kindt, T. J., Ewenstein, B. M., Uehara, H., Nisizawa, T. and Nathenson, S. G. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 3390–4.
Collins, J. H., Kramias, E. G., Reeves, A. S., Bilezikjian, L. M. and Schwartz, A. (1981) *Biochem. Biophys. Res. Commun.*, **99**, 796–803.
Corruzzi, G., Broglie, R., Cashmore, A. and Chua, N-H. (1983) *J. Biol. Chem.*, **258**, 1399–402.
Corruzzi, G. and Tzagoloff, A. (1979) *J. Biol. Chem.*, **254**, 9324–30.
Cosman, D., Gerretti, D. P., Larsen, A., Park, L., March, C., Dower, S., Gillis, S. and Urdel, D. (1984) *Nature (London)*, **312**, 768–71.
Cozens, A. L., Walker, J. E., Phillips, A. L., Huttly, A. K. and Gray, J. C. (1986) *EMBO J.*, **5**, 217–22.
Darlison, M. G. and Guest, J. R. (1984) *Biochem. J.*, **223**, 507–17.
Dawson, A. J., Jones, V. P. and Leaver, C. J. (1984) *EMBO J.*, **3**, 2107–13.
de Bruijn, M. H. L. (1983) *Nature (London)*, **304**, 234–41.
Dean, G. E., MacNab, R. M., Stader, J., Matsumara, P. and Burks, C. (1984) *J. Bacteriol.*, **159**, 991–9.
Deisenhofer, J., Epp, O., Miki, K., Huber, R. and Michel, H. (1984) *J. Mol. Biol.*, **180**, 385–98.
Devillers-Thiery, A., Giraudet, J., Bentaboulet, M. and Changeux, J-P. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 2067–71.
Drickamer, K., Mamon, J. F., Binns, G. and Leung, J. O. (1984) *J. Biol. Chem.*, **259**, 770–8.

Dunn, R., McCoy, J., Simsek, M., Majumdar, A., Chang, S. H., Rajbhandary, U. L. and Khorana, H. G. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 6744–8.
Edman, P. and Begg, G. (1967) *Eur. J. Biochem.*, **I**, 80–91.
Edman, P. and Henschen, A. (1975) in *Protein Sequence Determination* (ed S. B. Needleman), Springer-Verlag, Berlin, Heidelberg and New York, pp. 232–79.
Eisenberg, D. (1984) *Annu. Rev. Biochem.*, **53**, 595–623.
Ellingboe, J., Nyström, E. and Sjövall, J. (1969) *Methods Enzymol.*, **14**, 317–29.
Engelman, D. M., Goldman, A. and Steitz, T. A. (1982) *Methods Enzymol.*, **88**, 81–8.
Engelman, D. M., Henderson, R., McLachlan, A. D. and Wallace, B. A. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 2023–7.
Erdweg, M. and Buse, G. (1985) *Biol. Chem. Hoppe Seyler*, **366**, 257–63.
Ferguson, M. A. J., Haldar, K. and Cross, G. A. M. (1985) *J. Biol. Chem.*, **260**, 4963–8.
Fields, S., Winter, G. and Brownlee, G. G. (1981) *Nature (London)*, **290**, 213–17.
Fillingame, R. H. (1976) *J. Biol. Chem.*, **251**, 6630–7.
Finer-Moore, J. and Stroud, R. M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 155–9.
Folch, J. and Lees, M. (1951) *J. Biol. Chem.*, **191**, 807–17.
Foster, J. A., Bruenger, G. L., Hu, C. L., Albertson, K. and Franzblau, C. (1973) *Biochem. Biophys. Res. Commun.*, **53**, 70–4.
Fox, T. D. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 6534–8.
Fox, T. D. and Leaver, C. J. (1981) *Cell*, **26**, 315–23.
Frank, H., Chaves das Neves, H. J. and Bayer, E. (1978) *J. Chromatogr.*, **152**, 357–62.
Frank, H. and Desiderio, D. M. (1978) *Anal. Biochem.*, **90**, 413–19.
Frielle, T., Brunner, J. and Curthoys, N. P. (1982) *J. Biol. Chem.*, **257**, 14979–82.
Froshauer, S. and Beckwith, J. (1984) *J. Biol. Chem.*, **259**, 10896–903.
Garber, E. A., Krueger, J. G., Hanfusa, H. and Goldberg, A. R. (1983) *Nature (London)*, **302**, 161–3.
Garoff, H., Frischauf, A-M., Simons, K., Lehrach, H. and Delius, H. (1980) *Nature (London)*, **288**, 236–41.
Gay, N. J. and Walker, J. E. (1981) *Nucleic Acids Res.*, **9**, 3919–26.
Gay, N. J. and Walker, J. E. (1985) *EMBO J.*, **4**, 3519–24.
Gerber, G. E., Anderegg, R. J., Herlihy, W. C., Gray, C. P., Biemann, K. and Khorana, H. G. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 227–31.
Gething, M. J., Bye, J., Skehel, J. and Waterfield, M. (1980) *Nature (London)*, **287**, 301–6.
Gibson, B. W. and Biemann, K. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1956–60.
Graf, T. and Sebald, W. (1978) *FEBS Lett.*, **94**, 218–22.
Green, N. M. and Toms, E. J. (1985) *Biochem. J.*, **231**, 425–9.
Grisi, E., Brown, T. A., Waring, R. B., Scazzocchio, C. and Davies, R. W. (1982) *Nucleic Acids Res.*, **10**, 3531–9.
Grundström, T. and Jaurin, B. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 1111–15.
Haldar, K., Ferguson, M. A. J. and Cross, G. A. M. (1985) *J. Biol. Chem.*, **260**, 4969–74.
Hantke, K. and Braun, V. (1979) *Eur. J. Biochem.*, **34**, 284–96.
Hargrave, D. A., Fong, S. L., McDowell, J. H., Mas, M. T., Curtis, D. R., Wang, J. K., Siemiatkowski-Juszczak, E. and Smith, D. P. (1980) *Neurochem. Int.*, **1**, 231–44.
Hayes, P. K., Walsby, A. E. and Walker, J. E. (1986) *Biochem. J.*, **236**, 31–36.
Hedrick, S. M., Nielsen, E. A., Kavaler, J., Cohen, D. I. and Davis, M. M. (1984) *Nature (London)*, **308**, 153–8.
Heinemeyer, W., Alt, J. and Hermann, R. G. (1984) *Curr. Genet.*, **8**, 543–9.
Helenius, A. and Simons, K. (1975) Biochim. Biophys. Acta, **415**, 29–79.

Henderson, L. E., Krutzsch, H. C. and Oroszlan, S. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 339–43.

Henderson, R. (1981) in *Membranes and Intracellular Communication* (eds R. Balian *et al.*), North-Holland, Amsterdam, pp. 232–49.

Hesse, J. E., Wieczorek, L., Altendorf, K., Reicin, A. S., Dorus, E. and Epstein, W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4746–50.

Hewick, R. M., Hunkapiller, M. W., Hood, L. E. and Dreyer, W. J. (1981) *J. Biol. Chem.*, **256**, 7990–7.

Higgins, C. F., Haag, P. D., Nikaido, K., Ardeshiri, F., Garcia, G. and Gerro-Luzzi-Ames, G. (1982) *Nature (London)*, **298**, 723–7.

Higgins, R. C. and Dahmus, M. E. (1979) *Anal. Biochem.*, **93**, 257–60.

Hjelmel, L. M. and Chrambach, A. (1984) *Methods Enzymol.*, **104**, 305–47.

Holder, A. A., Lockyer, M. J., Odink, K. G., Sandhu, J. S., Riveros-Moreno, V., Nicholls, S. C., Hillman, Y., Davey, L. S., Tizard, M. L. V., Schwartz, R. T. and Freeman, R. R. (1985) *Nature*, **317**, 270–3.

Hoppe, J., Brunner, J. and Jorgensen, B. B. (1984) *Biochemistry*, **23**, 5610–16.

Hoppe, J. and Sebald, W. (1980) *Eur. J. Biochem.*, **107**, 57–65.

Horn, M. and Laursen, R. A. (1973) *FEBS Lett.*, **36**, 285–9.

Howe, C. J., Auffret, A. D., Doherty, A., Bowman, C. M., Dyer, T. A. and Gray, J. C. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6903–7.

Hudson, G. S., Farrell, P. J. and Barrell, B. G. (1985) *J. Virol.*, **53**, 528–35.

Hunkapiller, M. W. and Hood, L. E. (1980) *Science*, **207**, 523–5.

Hunkapiller, M. W., Lujan, E., Ostrander, F. and Hood, L. E. (1983a) *Methods Enzymol.*, **91**, 227–36.

Hunkapiller, M. W., Hewick, R. M., Dreyer, W. J. and Hood, L. E. (1983b) *Methods Enzymol.*, **91**, 399–13.

Inokuchi, K., Mutoh, N., Matsuyama, S. and Mizushima, S. (1982) *Nucleic Acids Res.*, **10**, 6957–68.

Jaiswal, A. K., Gonzalez, F. J. and Nebert, D. W. (1985) *Nucleic Acids Res.*, **13**, 4503–20.

Jollès, J., Schoentgen, F., Jollès, P., Vacher, M., Nicot, C. and Alfsen, A. (1979) *Biochem. Biophys. Res. Commun.*, **87**, 619–26.

Jorgensen, P. and Brunner, J. (1983) *Biochim. Biophys. Acta*, **735**, 291–6.

Kahan, I. and Moscarello, M. A. (1985) *Biochemistry*, **24**, 538–44.

Kaufman, J. F., Auffrey, C., Korman, A. J., Shackelford, D. A. and Strominger, J. (1984) *Cell*, **36**, 1–13.

Kawakami, K., Noguchi, S., Noda, M., Takahashi, H., Ohta, T., Kawamura, M., Nojima, H., Hirose, T., Inayama, S., Hayashida, H., Miyata, T. and Numa, S. (1985) *Nature (London)*, **316**, 733–6.

Kelley, J. A., Nau, N., Forster, H. J. and Biemann, K. (1975) *Biomed. Mass Spectrom.*, **2**, 313–25.

Khorana, H. G., Gerber, G. E., Herlihy, W. C., Gray, C. P., Anderegg, R. J., Nihei, K. and Biemann, K. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 5046–50.

Kimura, S., Gonzalez, F. J. and Nebert, D. W. (1984) *J. Biol. Chem.*, **259**, 10705–13.

Klapper, D. G., Wilde, C. E. and Capra, J. D. (1978) *Anal. Biochem.*, **85**, 126–31.

Kleffel, B., Garavito, R. M., Baumeister, W. and Rosenbusch, J. P. (1985) *EMBO J.*, **4**, 1589–92.

Kolata, G. (1985) *Science*, **229**, 850.

Kondo, K., Tajima, S., Sato, R. and Narita, K. (1979) *J. Biochem. (Tokyo)*, **86**, 1119–28.
Kopito, R. R. and Lodish, H. F. (1985) *Nature (London)*, **316**, 234–8.
Kubo, T., Noda, M., Takai, T., Tanabe, T., Kayano, T., Shimizu, S., Tanaka, K., Takahishi, H., Hirose, T., Inayama, S., Kikuno, R., Miyata, T. and Numa, S. (1985) *Eur. J. Biochem.*, **149**, 5–13.
Kyte, J. and Doolittle, R. F. (1982) *J. Mol. Biol.*, **157**, 105–32.
Laursen, R. A. (1971) *Eur. J. Biochem.*, **20**, 89–102.
Laursen, R. A., Bonner, A. G. and Horn, M. J. (1975) in *Instrumentation in Amino Acid Sequence Determination* (ed R. N. Perham), Academic Press, London, New York and San Francisco, pp. 73–110.
Laursen, R. A., Samiullah, M. and Lees, M. B. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2912–16.
Lees, M. B., Chao, B. H., Leu-Fen, H. L., Samiullah, M. and Laursen, R. A. (1983) *Arch. Biochem. Biophys.*, **226**, 643–56.
Lemke, G. and Axel, R. (1985) *Cell*, **40**, 501–8.
Macino, G. and Tzagoloff, A. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 131–5.
Macino, G. and Tzagoloff, A. (1980) *Cell*, **20**, 507–17.
MacLennan, D. H., Brandl, C. J., Korczak, B. and Green, N. M. (1985) *Nature (London)*, **316**, 696–700.
Macreadie, I. G., Novitski, C. E., Maxwell, R. J., John, U., Ooi, B. G., McMullen, G. L., Lukins, H. B., Linnane, A. W. and Nagley, P. (1983) *Nucleic Acids Res.*, **11**, 4435–51.
Maddy, A. H. (1964) *Biochim. Biophys. Acta*, **88**, 390–9.
Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor.
Mansfield, R. W. and Anderson, J. M. (1985) *Biochim. Biophys. Acta*, **809**, 435–44.
Martin, B., Svendsen, I. and Ottesen, M. (1977) *Carlsberg Res. Commun.*, **42**, 99–102.
Matsumoto, M., Matsumoto, R. and Folch-Pi, J. (1964) *J. Neurochem.*, **11**, 829–38.
Maxam, A. M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 560–64.
McConnell, D. G., Kohnker, R. E. and Smith, A. J. (1984) *Fed. Proc. Fed. Am. Soc. Exp. Biol.*, **43**, 1585.
Mihara, K. and Sato, R. (1985) *EMBO J.*, **4**, 769–74.
Mizuno, T., Chou, M-Y. and Inouye, M. (1983) *J. Biol. Chem.*, **258**, 6932–40.
Mokrasch, L. C. (1967) *Life Sci.*, **6**, 1905–9.
Morohashi, K., Fujii-Kuriyama, Y., Okada, Y., Sogana, K., Hirose, T., Inayama, S. and Omura, T. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4647–51.
Morris, H. R. and Dell, A. (1975) in *Amino Acid Sequence Determination* (ed R. N. Perham), Academic Press, London, New York and San Francisco, pp. 147–91.
Morris, H. R., Williams, D. H. and Ambler, R. P. (1971) *Biochem. J.*, **125**, 189–201.
Morris, J. and Herrmann, R. G. (1984) *Nucleic Acids Res.*, **12**, 2837–49.
Mostov, K. E., Friedlander, M. and Blobel, G. (1984) *Nature (London)*, **308**, 37–43.
Mueckler, M., Caruso, C., Baldwin, S. A., Panico, M., Blench, I., Morris, H. R., Allard, W. J., Lienhard, G. E. and Lodish, H. F. (1985) *Science*, **229**, 941–5.
Mullen, E. and Akhtar, M. (1981) *FEBS Lett.*, **132**, 261–4.
Nathans, J. and Hogness, D. S. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4851–5.
Nau, H. (1976) *Angew. Chem. (Int. Edn)*, **15**, 75–86.
Nau, H. and Biemann, K. (1976) *Anal. Biochem.*, **73**, 139–86.
Nelles, L. P. and Bamburg, J. R. (1976) *Anal. Biochem.*, **73**, 522–31.

Nelson, D. R. and Robinson, N. C. (1983) *Methods Enzymol.*, **97**, 571–618.
Nobrega, F. G. and Tzagoloff, A. (1980) *J. Biol. Chem.*, **255**, 9828–37.
Noda, M., Furutani, Y., Takahashi, H., Toyosato, M., Tanabe, T., Shimizu, S., Kikyotani, S., Kayano, T., Hirose, T., Inayama, S. and Numa, S. (1983c) *Nature (London)*, **305**, 818–23.
Noda, M., Shimizu, S., Tanabe, T., Takei, T., Kayano, T., Ikeda, T., Takahashi, H., Nakayama, H., Kanaoka, Y., Minamino, N., Kangawa, K., Matsuo, H., Raftery, M. A., Hirose, T., Inayama, S., Hayashida, H., Miyata, T. and Numa, S. (1984) *Nature (London)*, **312**, 121–7.
Noda, M., Takahashi, H., Tanabe, T., Toyosato, M., Furutani, Y., Hirose, T., Asai, M., Inayama, S., Miyata, T. and Numa, S. (1982) *Nature (London)*, **299**, 793–7.
Noda, M., Takahashi, H., Tanabe, T., Toyosato, M., Kikyotani, S., Furutani, Y., Hirose, T., Takashima, H., Inayama, S., Miyata, T. and Numa, S. (1983b) *Nature (London)*, **302**, 528–31.
Noda, M., Takahashi, H., Tanabe, T., Toyosato, M., Kikyotani, S., Hirose, T., Asai, M., Takashima, H., Inayama, S., Miyata, T. and Numa, S. (1983a) *Nature (London)*, **301**, 251–5.
Omary, M. B. and Trowbridge, I. S. (1981) *J. Biol. Chem.*, **256**, 4715–18.
Ovchinnikov, Yu. A., Abdulaev, N. G., Feigina, M. Yu., Artamonov, I. D., Bogachuk, A. S., Zolotarev, A. S., Eganyan, E. R. and Kostetsky, P. V. (1983) *Bioorg. Khim.*, **9**, 1331–40.
Ovchinnikov, Yu. A., Abdulaev, N. G., Feigina, M. Yu, Kiselev, A. V. and Lobanov, N. A. (1979) *FEBS Lett.*, **100**, 219–24.
Overbeeke, N., Bergmans, H., van Mansfield, F. and Lugtenberg, B. (1983) *J. Mol. Biol.*, **163**, 513–32.
Ozols, J., Carr, S. A. and Strittmatter, P. (1984) *J. Biol. Chem.*, **259**, 13349–54.
Ozols, J. and Heinemann, F. S. (1982) *Biochim. Biophys. Acta*, **704**, 163–73.
Paul, C. and Rosenbusch, J. P. (1985) *EMBO J.*, **4**, 1593–7.
Pellman, D., Garber, E. A., Cross, F. R. and Hanafusa, H. (1985) *Nature (London)*, **314**, 374–7.
Phillips, A. L. and Gray, J. C. (1984) *Mol. Gen. Genet.*, **194**, 477–84.
Ploegh, H. L., Orr, H. T. and Strominger, J. (1981) *Cell*, **24**, 287–99.
Pober, J. S. and Stryer, L. (1975) *J. Mol. Biol.*, **95**, 477–81.
Porter, A. G., Barber, C., Carey, N. H., Hallewell, R. A., Threlfall, G. and Emtage, J. S. (1979) *Nature (London)*, **282**, 471–7.
Reeves, A. S., Collins, J. H. and Schwartz, A. (1980) *Biochem. Biophys. Res. Commun.*, **95**, 1591–8.
Rifkin, D. B., Compans, R. W. and Reich, E. (1972) *J. Biol. Chem.*, **247**, 6432–7.
Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L. and Wall, R. (1980) *Cell*, **20**, 303–12.
Rose, G. D., Gierasch, L. M. and Smith, J. A. (1985) *Adv. Protein Chem.*, **37**, 1–109.
Rose, G. D., Young, W. B. and Gierasch, L. M. (1983) *Nature (London)* **304**, 654–7.
Rose, J. K., Welch, W. J., Sefton, B. M., Esch, F. S. and Ling, N. C. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 3884–8.
Rouser, G., Kritchevsky, G., Yamamoto, A., Simon, G., Galli, L. and Bauman, A. J. (1969) *Methods Enzymol.*, **14**, 272–317.
Rubinstein, M. (1979) *Anal. Biochem.*, **98**, 1–7.

Rubinstein, M., Stein, S. and Udenfriend, S. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 669–71.
Sadler, I., Sadu, K., Schatz, G., Kaudewitz, F. and Haid, A. (1984) *EMBO J.*, **3**, 2137–43.
Sale, G. J., Towner, P. and Akhtar, M. (1977) *Biochemistry*, **16**, 5641–8.
Sanderman, H. and Strominger, J. L. (1972) *J. Biol. Chem.*, **247**, 5123–31.
Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–7.
Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., Hutchinson III, C. A., Slocombe, P. M. and Smith, M. (1979) *Nature (London)*, **265**, 687–95.
Saraste, M. (1984) *FEBS Lett.*, **166**, 367–72.
Saraste, M. and Walker, J. E. (1982) *FEBS Lett.*, **144**, 250–4.
Savige, W. E. and Fontana, A. (1977) *Methods Enzymol.*, **47**, 459–69.
Schiffer, M. and Edmundson, A. B. (1967) *Biophys. J.*, **7**, 121–35.
Schneider, C., Owen, M. J., Banville, D. and Williams, J. G. (1984) *Nature (London)*, **311**, 675–8.
Schultz, A. M., Henderson, L. E., Oroszlan, S., Garber, E. A. and Hanfusa, S. (1985) *Science*, **227**, 427–9.
Schultz, A. M. and Oroszlan, S. (1984) *Virology*, **133**, 431–7.
Sebald, W. and Hoppe, J. (1981) *Curr. Top. Bioenerget.*, **12**, 1–64.
Sebald, W., Graf, T. and Lukins, H. B. (1979) *Eur. J. Biochem.*, **93**, 587–99.
Sefton, B. M., Trowbridge, I. S., Cooper, J. A. and Scolnick, E. M. (1982) *Cell*, **31**, 465–74.
Segrest, J. P. and Feldmann, R. J. (1974) *J. Mol. Biol.*, **87**, 853–8.
Segrest, J. P., Jackson, R. L., Marchesi, V. T., Guyer, R. B. and Terry, W. (1972) *Biochem. Biophys. Res. Commun.*, **49**, 964–9.
Seki, T., Chang, H. C., Moriuchi, T., Denome, T., Ploegh, H. and Silver, J. (1985) *Science*, **227**, 649–51.
Shaw, D. C., Walker, J. E., Northrop, F. D., Barrell, B. G., Godson, G. N. and Fiddes, J. C. (1978) *Nature (London)*, **272**, 510–15.
Shibahara, S., Kubo, T., Perski, H. J. Takahashi, H., Noda, M. and Numa, S. (1985) *Eur. J. Biochem.*, **146**, 15–22.
Shull, G. E., Schwartz, A. and Lingrel, J. B. (1985) *Nature (London)*, **316**, 691–5.
Sim, G. K., Vagüe, J., Nelson, J., Marrack, P., Palmer, E., Augustin, A. and Kappler, J. (1984) *Nature (London)*, **312**, 771–5.
Smith, M. (1983) in *Methods of DNA and RNA Sequencing* (ed S. M. Weismann), Praeger, New York, pp. 23–68.
Soto, E. F., Pasquiri, J. M., Placido, R. and La Torre, J. L. (1969) *J. Chromatogr.*, **41**, 400–9.
Spiess, M., Brunner, J. and Semenza, G. (1982) *J. Biol. Chem.*, **257**, 2370–7.
Spiess, M. and Lodish, H. F. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6465–9.
Spiess, M., Schwartz, A. L. and Lodish, H. F. (1985) *J. Biol. Chem.*, **260**, 1979–82.
Stearne, P. A., Pietersz, G. A. and Goding, J. W. (1985) *J. Immunol.*, **134**, 3474–9.
Steffens, G. and Buse, G. (1976) *Hoppe-Seyler's Z. Physiol. Chem.*, **357**, 1125–37.
Steffens, G. J. and Buse, G. (1979) *Hoppe-Seyler's Z. Physiol. Chem.*, **360**, 613–19.
Steffens, K., Kiltz, H. H., Schneider, E., Schmid, R. and Altendorf, K. (1982) *FEBS Lett.*, **142**, 151–4.

Stieger, S., Brodbeck, U., Reber, B. and Brunner, J. (1984) *FEBS Lett.*, **168**, 231–4.
Stoffel, W., Giersiefen, H., Hillen, H., Schroeder, W. and Tunggaal, B. (1985) *Biol. Chem. Hoppe-Seyler*, **366**, 627–35.
Stoffel, W., Hillen, H. and Giersiefen (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5012–16.
Stoffel, W., Hillen, H., Schröder, W. and Deutzmann, R. (1982a) *Hoppe-Seyler's Z. Physiol. Chem.*, **363**, 855–64.
Stoffel, W., Hillen, H., Schröder, W. and Deutzmann, R. (1983) *Hoppe Seyler's Z. Physiol. Chem.*, **364**, 1455–66.
Stoffel, W., Schröder, W., Hillen, H. and Deutzmann (1982b) *Hoppe Seyler's Z. Physiol. Chem.*, **363**, 1117–31.
Strickler, J. E., Hunkapiller, M. W. and Wilson, K. J. (1984) *Anal. Biochem.*, **140**, 553–66.
Sudhoff, T. C., Goldstein, J. L., Brown, M. S. and Russell, D. W. (1985) *Science*, **288**, 815–22.
Sumikawa, K., Houghton, M., Smith, J. C., Richards, B. M. and Barnard, E. A. (1982) *Nucleic Acids Res.*, **10**, 5809–12.
Surin, B. P., Jans, D. A., Fimmel, A. L., Shaw, D. C., Cox, G. B. and Rosenburg, H. (1984) *J. Bacteriol.*, **157**, 772–8.
Sutton, M. R., Rosen, D., Feher, G. and Steiner, L. A. (1982) *Biochemistry*, **21**, 3842–9.
Takagaki, Y., Gerber, G. E., Nihei, K. and Khorana, H. G. (1980) *J. Biol. Chem.*, **255**, 1536–41.
Takagaki, Y., Radhakrishnan, R., Gupta, C. M. and Khorana, H. G. (1983) *J. Biol. Chem.*, **258**, 9128–35.
Takahashi, Y., Kato, K., Hayashizaki, Y., Wakabayashi, T., Ohtsuka, E., Matsuki, S., Ikehara, M. and Matsubara, K. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1931–5.
Takei, T., Noda, M., Furutani, Y., Takahashi, H., Notaka, M., Shimizu, S., Kayano, T., Tanabe, T., Tanaka, K., Hirose, T., Inayama, S. and Numa, S. (1984) *Eur. J. Biochem.*, **143**, 109–15.
Takei, T., Noda, M., Mishina, M., Shimizu, S., Furutani, Y., Kayano, T., Ikeda, T., Kubo, T., Takahashi, H., Takahashi, T., Kuno, M. and Numa, S. (1985) *Nature*, **315**, 761–4.
Talbot, D. N. and Yphantis, D. A. (1971) *Anal. Biochem.*, **44**, 246–53.
Tanabe, T., Noda, M., Furutani, Y., Takei, T., Takahashi, H., Tanaka, K., Hirose, T., Inayama, S. and Numa, S. (1984) *Eur. J. Biochem.*, **144**, 11–17.
Tanabe, T., Nukada, T., Nishikawa, Y., Sugimoto, K., Suzuki, H., Takahashi, H., Noda, M., Hagg, T., Ichiyama, A., Kangawa, K., Minamino, N., Matsuo, H. and Numa, S. (1985) *Nature (London)*, **315**, 242–5.
Tanaka, M., Hanio, M., Yasunobo, K. T., Yu, C. A., Yu, L., Wei, Y-H. and King, T. E. (1979) *J. Biol. Chem.*, **254**, 3879–85.
Tanford, C. and Reynolds, J. A. (1976) *Biochim. Biophys. Acta*, **457**, 132–70.
Tarr, G., Beecher, J. F., Bell, M. and McKean, D. J. (1978) *Anal. Biochem.*, **84**, 622–7.
Tarr, G. E. and Crabb, J. W. (1983) *Anal. Biochem.*, **131**, 99–107.
Thalenfield, B. E. and Tzagoloff, A. (1980) *J. Biol. Chem.*, **255**, 6173–80.
Theiler, R., Suter, F. and Zuber, H. (1983) *Hoppe-Seyler's Z. Physiol. Chem.*, **364**, 1765–76.
Theiler, T., Suter, F., Wiemken, V. and Zuber, H. (1984) *Hoppe-Seyler's Z. Physiol. Chem.*, **365**, 703–19.

Tomita, M. and Marchesi, V. T. (1975) *Proc. Natl. Acad. Sci. USA*, **72**, 2964–8.
Tonn, S. J., Gogel, G. E. and Loach, P. A. (1977) *Biochemistry*, **16**, 877–85.
Towner, P., Sale, G. J. and Akhtar, M. (1977) *FEBS Lett.*, **76**, 51–5.
Tschesche, H. and Kupfer, S. (1972) *Eur. J. Biochem.*, **26**, 33–6.
Ullrich, A., Bell, J. R., Chen, E. Y., Herrera, T., Petruzzelli, L. M., Dull, T. J., Gray, A., Coussens, L., Liao, Y.-C., Tsubokawa, M., Mason, A., Seeburg, P. H., Grunfeld, C., Rosen, O. M. and Ramachandran, J. (1985) *Nature (London)*, **313**, 756–61.
Ullrich, A., Coussens, L., Hayflick, J. S., Dull, T. J., Gray, A., Tam, A. W., Lee, J., Yarden, Y., Liberman, T. A., Schlessinger, J., Downward, J., Mayes, E. L. V., Whittle, N., Waterfield, M. D. and Seeburg, P. H. (1984) *Nature (London)*, **309**, 418–25.
Unwin, N. and Henderson, R. (1984) *Sci. Am.*, **250**, 56–66.
Velours, J., Esparza, M. and Guerin, B. (1982) *Biochem. Biophys. Res. Commun.*, **109**, 1192–9.
Velours, J., Esparza, M., Hoppe, J., Sebald, W. and Guerin, B. (1984) *EMBO J.*, **3**, 207–12.
Verhoeyen, M., Fang, R., Jou, W. H., Deuos, R., Huylebroeck, D., Saman, E. and Fiers, W. (1980) *Nature (London)*, **286**, 771–6.
Viebrok, A., Perz, A. and Sebald, W. (1982) *EMBO J.*, **1**, 565–71.
Wachter, E., Hofner, H. and Macleidt, W. (1975) in *Solid Phase Methods in Protein Sequence Analysis* (ed R. A. Laursen), Pierce Chemical Co., Rockford, Illinois, pp. 31–46.
Wachter, E. and Werhahn, R. (1979) *Anal. Biochem.*, **97**, 56–64.
Wakabayashi, S., Matsubara, H., Kim, C. H., Kawa, K. and King, T. S. (1980) *Biochem. Biophys. Res. Commun.*, **97**, 1548–54.
Wakabayashi, S., Matsubara, H., Kim, C. H. and King, T. E. (1982) *J. Biol. Chem.*, **257**, 9335–44.
Walker, J. E., Auffret, A. D., Carne, A., Gurnett, A., Hanisch, P., Hill, D. and Saraste, M. (1982b) *Eur. J. Biochem.*, **123**, 253–60.
Walker, J. E., Auffret, A. D., Carne, A. F., Naughton, M. A. and Runswick, M. J. (1980) in *Methods in Peptide and Protein Sequence Analysis* (ed Ch. Birr), Elsevier/North-Holland Biomedical Press, Amsterdam, pp. 257–65.
Walker, J. E., Carne, A. F. and Schmitt, H. W. (1979) *Nature (London)* **278**, 653–4.
Walker, J. E., Eberle, A., Gay, N. J., Hanisch, P., Saraste, M. and Runswick, M. J. (1982a) in *Methods in Peptide and Protein Sequence Analysis* (ed M. Elzinga), Humana, Clifton, New Jersey, pp. 337–54.
Walker, J. E., Fearnley, I. M., Gay, N. J., Gibson, B. W., Northrop, F. D., Powell, S. J., Runswick, M. J., Saraste, M. and Tybulewicz, V. L. J. (1985) *J. Mol. Biol.*, **184**, 677–701.
Waring, R. B., Brown, T. A., Ray, J. A., Scazzocchio, C. and Davies, R. W. (1984) *EMBO J.*, **3**, 2121–8.
Waring, R. B., Davies, R. W., Lee, S., Grisi, E., Berks, M. and Scazzocchio, C. (1981) *Cell*, **27**, 4–11.
Wechsler, T., Suter, F., Fuller, C. and Zuber, H. (1985) *FEBS Lett.*, **181**, 173–8.
Wessel, D. and Flügge, U. I. (1984) *Anal. Biochem.*, **138**, 141–3.
Widger, W. R., Cramer, W. A., Herrman, R. G. and Trebst, A. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 674–8.
Willey, D. L., Auffret, A. D. and Gray, J. C. (1984a) *Cell*, **36**, 555–62.

Willey, D. W., Howe, C. H., Auffret, A. D., Bowman, C. M., Dyer, T. A. and Gray, J. C. (1984b) *Mol. Gen. Genet.*, **194**, 416–22.

Winter, G., Fields, S. and Brownlee, G. G. (1981) *Nature (London)*, **292**, 72–5.

Wood, D., Darlison, M. G., Wilde, R. J. and Guest, J. R. (1984) *Biochem. J.*, **222**, 519–34.

Wood, W. I., Gitschier, J., Lasky, L. A. and Lawn, R. M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1585–8.

Yabusaki, Y., Shimizu, M., Murakami, H., Nakamura, K., Oeda, K. and Ohkawa, H. (1984) *Nucleic Acids Res.*, **12**, 2929–38.

Yanagi, Y., Yoshikai, Y., Leggett, K., Clark, S. P., Aleksander, I. and Mak, T. W. (1984) *Nature (London)*, **308**, 145–9.

Young, I. G., Rogers, B. L., Campbell, H. D., Jaworowski, A. and Shaw, D. C. (1981) *Eur. J. Biochem.*, **116**, 165–70.

Youvan, D. C., Bylina, E. J., Alberti, M., Begusch, H. and Hearst, J. E. (1984) *Cell*, **37**, 949–57.

Zulauf, M. and Rosenbusch, J. P. (1983) *J. Phys. Chem.*, **87**, 856–62.

Zurawski, G., Bohnert, H. J., Whitfeld, P. R. and Bottomley, W. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 7699–703.