Biochem. J. (1974) **141**, 243-255
Printed in Great Britain

Amino Acid Sequence of the Smaller Basic Protein from Rat Brain Myelin

By PETER R. DUNKLEY and PATRICK R. CARNEGIE
The Russell Grimwade School of Biochemistry, University of Melbourne, Parkville, Vic. 3052, *Australia*

(*Received* 27 *November* 1973)

1. The complete amino acid sequence of the smaller basic protein from rat brain myelin was determined. This protein differs from myelin basic proteins of other species in having a deletion of a polypeptide of 40 amino acid residues from the centre of the molecule. 2. A detailed comparison is made of the constant and variable regions in a group of myelin basic proteins from six species. 3. An arginine residue in the rat protein was found to be partially methylated. The ratio of methylated to unmethylated arginine at this position differed from that found for the human basic protein. 4. Three tryptic peptides were isolated in more than one form. The differences between the two forms of each peptide are discussed in relation to the electrophoretic heterogeneity of myelin basic proteins, which is known to occur at alkaline pH values. 5. Detailed evidence for the amino acid sequence of the protein has been deposited as Supplementary Publication SUP 50029 at the British Library (Lending Division) (formerly the National Lending Library for Science and Technology), Boston Spa, Yorks. LS23 7BQ, U.K., from whom copies may be obtained on the terms given in *Biochem. J.* (1973) **131**, 5.

Myelin from the central nervous system contains a basic protein which comprises 30% of the total myelin protein (Eng *et al.*, 1968) and which can induce experimental autoimmune encephalomyelitis when injected into laboratory animals (Alvord, 1968). The amino acid sequences of the human (Carnegie, 1971*a*) and bovine (Eylar *et al.*, 1971) basic proteins have been determined.

Rat central nervous system myelin contains two basic proteins (Mehl, 1967; Cotman & Mahler, 1967; Eng *et al.*, 1968; Martenson *et al.*, 1971*a*). Rodents of the suborders Myomorpha and Sciuromorpha have two myelin basic proteins (Martenson *et al.*, 1970*b*), whereas other species have a single protein comparable in amino acid composition, size (approx. 18400daltons) and electrophoretic mobility with the larger basic protein (L protein) from rats (Martenson *et al.*, 1970*a*). The second rat basic protein (S protein) is smaller by approximately 4000daltons (Martenson *et al.*, 1970*a*). Preliminary studies indicated that the S protein differed from the L protein in having 40 amino acid residues missing within the *C*-terminal half of the molecule (Dunkley *et al.*, 1972; Martenson *et al.*, 1972*b*). The partial sequence of an encephalitogenic peptide from the S protein has been reported (Dunkley *et al.*, 1973; McFarlin *et al.*, 1973).

The present paper gives evidence for the complete amino acid sequence of the rat S protein. Amino acid residues modified by methylation or phosphorylation are discussed. In this paper advantage has been taken of the data deposition scheme organized by the National Lending Library for Science and Technology to make full details of the composition and properties of peptides accessible to anyone interested, and thus to include only sufficient detail in the published paper to allow readers to assess the approach used. This work was presented in part at the 4th Meeting of the International Society for Neurochemistry, August, 1973 (Carnegie & Dunkley, 1973).

Experimental

Materials

Proteolytic enzymes used in this study were obtained from sources described by Carnegie (1971*a*). Synthetic methylated arginines were kindly supplied by Professor Y. Kakimoto, Osaka University Medical School, Osaka, Japan. Cytochrome *c* (horse heart, grade A) was obtained from Calbiochem, San Diego, Calif., U.S.A. Chromatography papers (3MM and no. 1) were Whatman products. Sephadex G-75, Sephadex G-150 and CM-Sephadex C-25 were obtained from Pharmacia Fine Chemicals, Uppsala, Sweden.

Isolation of rat S protein

The S protein was isolated from adult Buffalo rat brains essentially by the method of Dunkley & Carnegie (1974) and Deibler *et al.* (1972). Briefly this method involved defatting the fresh tissue with chloroform–methanol (2:1, v/v) (19ml/g), water and 5% (w/v) NaCl washes of the residue and extraction of

the proteins by homogenization in HCl (pH3.0, 2ml/g of original tissue). The rat basic proteins were concentrated and purified by batch adsorption on to CM-Sephadex C-25 (1g/20g of original tissue) at pH4.5. After thorough washing of the CM-Sephadex with water, it was poured into a column (1.8cm× 35cm) and the proteins were eluted with 0.2M-HCl. The protein peak eluted with the acid front was dialysed, freeze-dried and a sample (150mg) applied to a Sephadex G-75 column (3.3cm×130cm) equilibrated with 0.01M-HCl. The two rat basic proteins were eluted between 1.3 and 1.7 times the exclusion volume of this column. After dialysis and freeze-drying this fraction was applied to a Sephadex G-50 column (5.0cm×140cm) equilibrated in 0.2M-KCl–HCl buffer, pH2.0, to separate rat L and S proteins. Rechromatography of these fractions on Sephadex G-150 in 0.1M-HCl yielded S protein, which on electrophoresis at pH2.0 (Carnegie *et al.*, 1967) showed over 95% of the protein in a single band with a mobility of 0.9 relative to cytochrome *c*, when examined on a Zeiss PMQ11 spectrophotometer equipped with a scanning attachment and a 100μm slit device (Brinkman Instruments Inc.).

Amino acid analysis

Amino acid analyses of S protein and peptides were performed as described by Carnegie (1971*a*). Analysis of the protein before and after performic acid oxidation (Hirs, 1956) was used to assess methionine content, and tryptophan was determined by the method of Gaitonde & Dovey (1970). Methylarginine analyses, based on elution profiles of standard derivatives, were all performed under routine amino acid analysis conditions by using Technicon C-2 resin (batch no. 1-2-013-15).

Enzymic and chemical digestion of the rat S protein

(a) Trypsin (400μg) was incubated with S protein (40mg) in 1% NH_4HCO_3 for 3h at 37°C. (b) Pepsin (300μg) was incubated with S protein (20mg) in 5% (v/v) formic acid for 3h at 37°C. (c) Thermolysin (300μg) was incubated with S protein (30mg) in 1% NH_4HCO_3 for 30min at 37°C. (d) The S protein was digested with CNBr as described by Carnegie (1969).

Peptide isolations

Peptides obtained from the tryptic, thermolytic, peptic and CNBr digests of the S protein were designated with the prefixes T, L, P and CNBr respectively. The peptides were purified on paper by electrophoresis at pH6.5 and chromatography in butan-1-ol–acetic acid–water–pyridine (BAWP; 15:3:12:10, by vol.) (Carnegie, 1971*a*). In all cases peptides were purified to homogeneity as judged by the presence of a single ninhydrin-positive spot after paper electrophoresis at pH3.5, and by amino acid analysis in which the relative amounts of amino acids were close to integral values and contamination greater than 0.2 was considered unacceptable. In such cases isolation and analyses were repeated until satisfactory. For certain valine and isoleucine residues values as low as 0.6 were accepted, subject to the repeat hydrolysis for a prolonged period (120h) yielding integral values for these amino acids. Values of 0.6 were accepted for methionine and on this basis two residues were found, this result being in agreement with the accurate values obtained on analysis of the S protein. The amino acids present in the S protein were fully accounted for by the amino acids found in the major tryptic peptides, which were all in yields consistent with their having been derived from the S protein (Supplementary Publication SUP 50029). Slightly greater contamination was tolerated for a few of the larger overlap peptides. Where such contamination occurred the overlap peptides were digested with trypsin and the identity of the products was checked by peptide 'mapping' and amino acid analysis.

Specific staining of peptides containing tryptophan, methionine and histidine and/or tyrosine on paper was performed as described by Easley (1965). Peptides containing methylarginine residues were detected by amino acid analysis.

Mobilities of the peptides at pH6.5 (*m*) and in BAWP (R_{Leu}) were calculated as described by Carnegie (1971*a*).

Assignment of amide residues

Amide residues were assigned on the basis of electrophoretic mobilities of the peptides at pH6.5 (Offord, 1966) and supported by digestion with carboxypeptidases A and B or aminopeptidase O. The colour of certain peptides after cadmium–ninhydrin staining was also used in determining amide residues at *N*-terminal positions (Shotton & Hartley, 1973).

Sequence strategy

The sequence of the rat S protein was completed by isolating and sequencing all of the tryptic peptides and obtaining overlaps for these by digestion of the protein with pepsin, thermolysin and CNBr. Peptide sequences were determined by the dansyl–Edman procedure (Hartley, 1970). Where *N*-terminal amino acids were not clearly determined by this procedure, subtractive Edman analysis was performed. Certain peptides required further enzymic digestion by thermolysin, chymotrypsin, aminopeptidase O and carboxypeptidases A and B (Carnegie, 1971*a*). Except where indicated the

methods described above enabled positions of amino acid residues to be determined without reference to published sequences of myelin basic proteins from other species.

Results

Isolation of the rat S protein

Although the use of myelin as the starting material for extraction yielded less complex mixtures of protein for fractionation, it was not satisfactory for providing large quantities of basic protein. Thus whole brain was used and after defatting and acid extraction most contaminating proteins were separated by the CM-Sephadex procedure and subsequent gel filtration on Sephadex G-75. Gel filtration of myelin basic proteins was earlier performed in 0.01M-HCl (Deibler *et al.*, 1970). However, by increasing the concentration of the eluting solution to 0.1M-HCl (Dunkley & Carnegie, 1974) or by using 0.2M-KCl–HCl buffer (pH2.0), an improved separation of the rat L and S proteins from contaminants was achieved. Presumably this was due to a decreased interaction of the basic protein with the small number of carboxyl groups on Sephadex (Martenson *et al.*, 1970*a*). Long exposure of the basic protein to strong acid should be avoided. By using a large Sephadex G-50 column, in the KCl–HCl buffer the L and S proteins were rapidly separated, but rechromatography on Sephadex G-150 was required to obtain protein of sufficient purity. This was especially so for the L protein, which co-chromatographed with a brain protein of entirely different amino acid composition. This contaminant was not the histone F1 referred to by Martenson *et al.* (1971*a*) but could be their protein 'X'. The final yield of rat S protein was approx. 0.25mg/g of fresh tissue and the yield of L protein was fourfold less.

Amino acid sequences of tryptic peptides

A peptide 'map' of a tryptic digest of the rat S protein showed 24 major spots and a number of minor spots with the cadmium–ninhydrin stain. Specific stains indicated the presence of one peptide with tryptophan, two with methionine and seven with histidine and/or tyrosine. Analysis of the rat S protein gave 0.74mol of tryptophan and 1.6mol of methionine per mol of protein. Amino acid compositions of the 24 major tryptic peptides are given in the Supplementary Publication SUP 50029, and amino acid sequences of these peptides are shown in Fig. 1. The final yield of each peptide was between 5 and 15% of that expected from amino acid analysis of the S protein before digestion, and no corrections were made for losses on edge strips etc. Detailed properties (yields, electrophoretic properties, chromatographic properties and assignment of amide residues) for each of the tryptic peptides are discussed in full in the Supplementary Publication SUP 50029. For convenience trypic peptides are numbered according to their position in the final sequence starting from the *N*-terminal end. Only the more complex peptides are discussed below and no difficulties were experienced with the other tryptic peptides.

Peptide T1 was slow to react with ninhydrin and no *N*-terminal amino acid was detected by the dansyl procedure. A series of carboxypeptidase A and B digests of peptide T1 suggested the sequence -Ser-Gln-Lys. As the peptide was neutral at pH6.5 the *N*-terminal alanine residue was assumed to be blocked. Myelin basic proteins from human (Carnegie, 1971*a*) and bovine (Eylar *et al.*, 1971) sources have *N*-terminal acetylalanine residues and it was assumed that this was also the case with the S protein.

Two peptides were found with the same amino acid composition as peptide T4 (Fig. 1) but they had different chromatographic properties in BAWP. The slower-moving peptide (R_{Leu} 0.68) was always found in low yield and differed from peptide T4 (R_{Leu} 0.78) only in the oxidation state of its methionine residue (Harris & Roos, 1959). Thermolysin digestion of peptide T4 yielded free tyrosine and six peptides, T4L1–T4L6 (Fig. 1). A peptide L2 (Fig. 2), with identical mobility at pH6.5 and amino acid analysis with peptide T4L2, was obtained from a thermolysin digest of the whole S protein. Chymotryptic digestion of peptide T4 yielded, among others, a peptide T4C1 (Fig. 1). The peptide's mobility (*m* 0.22) at pH6.5 and red colour with cadmium–ninhydrin was consistent with it containing aspartic acid, not asparagine. The slightly acidic peptide T4L6 (*m* −0.06) confirmed this assignment. The amino acid composition and sequence of these thermolytic and chymotryptic peptides determined the amino acid sequence of peptide T4 (Fig. 1).

Digestion of peptide T5 with thermolysin yielded two peptides T5L1 and T5L2 whose amino acid composition and sequence established the sequence of the parent peptide.

Peptide T7 was partially sequenced by the dansyl–Edman procedure (Fig. 1). The composition and sequence of the thermolytic peptide L5 (Fig. 2), and the overlap peptides L6 and P3, provided evidence for the amino acid sequence in this region of the protein (Fig. 2).

Peptide T8 was sequenced by the dansyl–Edman procedure and the results were confirmed by subtractive Edman analysis because of disagreement (see the Discussion section) on the order of the serine and glycine residues (McFarlin *et al.*, 1973).

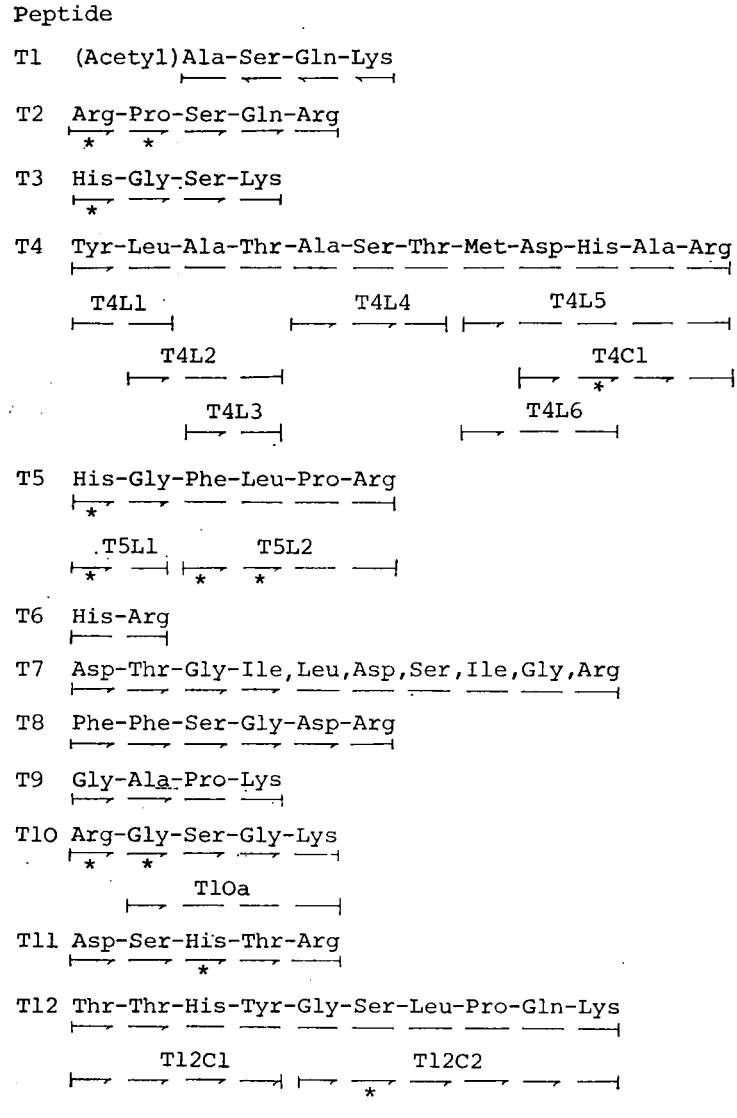
Peptide T10 presented minor problems. The low yield of peptide T10a, which was derived from peptide T10, was due to the slow rate of release of the *N*-terminal arginine residue by trypsin (Hill, 1965). Peptide T10 was also isolated joined to peptide

T11 as a result of slow hydrolysis of the Lys-Asp bond.

A second peptide was found with the same amino acid composition as peptide T11 but with different mobility (m −0.35) at pH6.5. Peptide T11 (m 0.34) contained an aspartic acid residue rather than asparagine. The difference in mobility indicated an addition to peptide T11 of two negative charges. The yield of this highly acidic peptide was approximately one-seventh that of peptide T11, and its possible origin is discussed below.

Digestion of peptide T12 with chymotrypsin yielded two peptides T12C1 and T12C2 whose amino acid composition and sequence determined the amino acid sequence of the parent peptide.

Digestion of peptide T14 with thermolysin yielded two peptides T14L1 and T14L2. Peptide T14L1 yielded only one valine residue by the usual amino acid-analysis procedure (22h hydrolysis), but hydrolysis for 120h showed a second valine residue; a Val-Val sequence was confirmed by peptide L9 (Fig. 2). Peptide T14 (m −0.22) was slightly acidic at pH6.5, suggesting the presence of two acidic amino acid residues. Peptide T14L1 was subjected to repeated Edman degradations and electrophoresis of the products was used to assign the acidic and amide



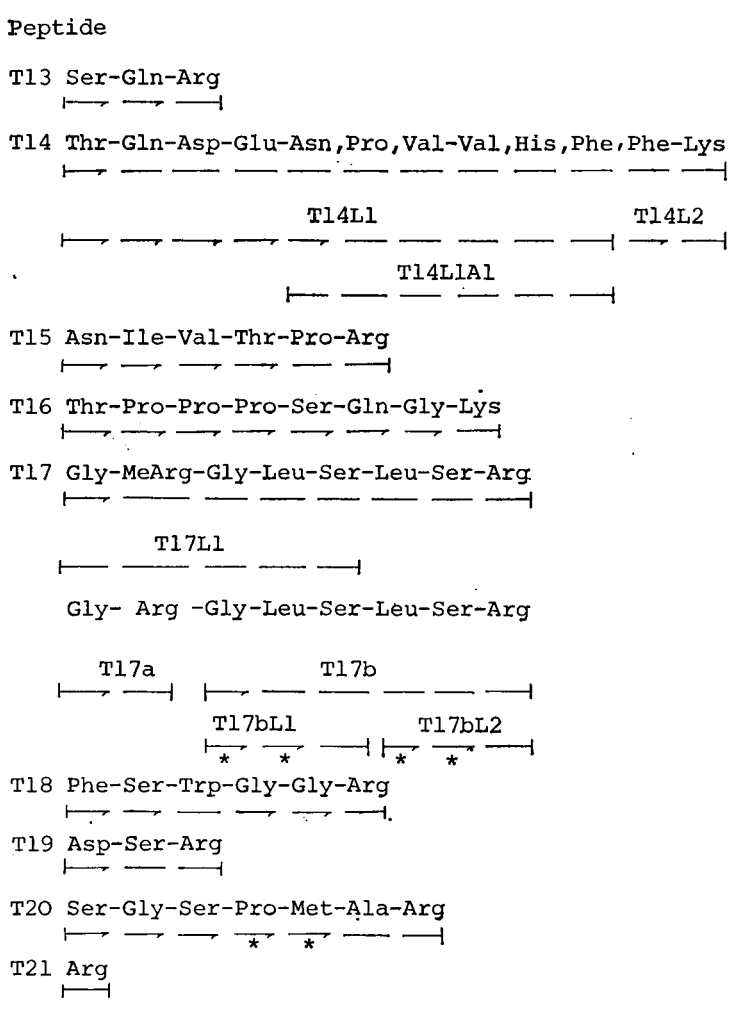


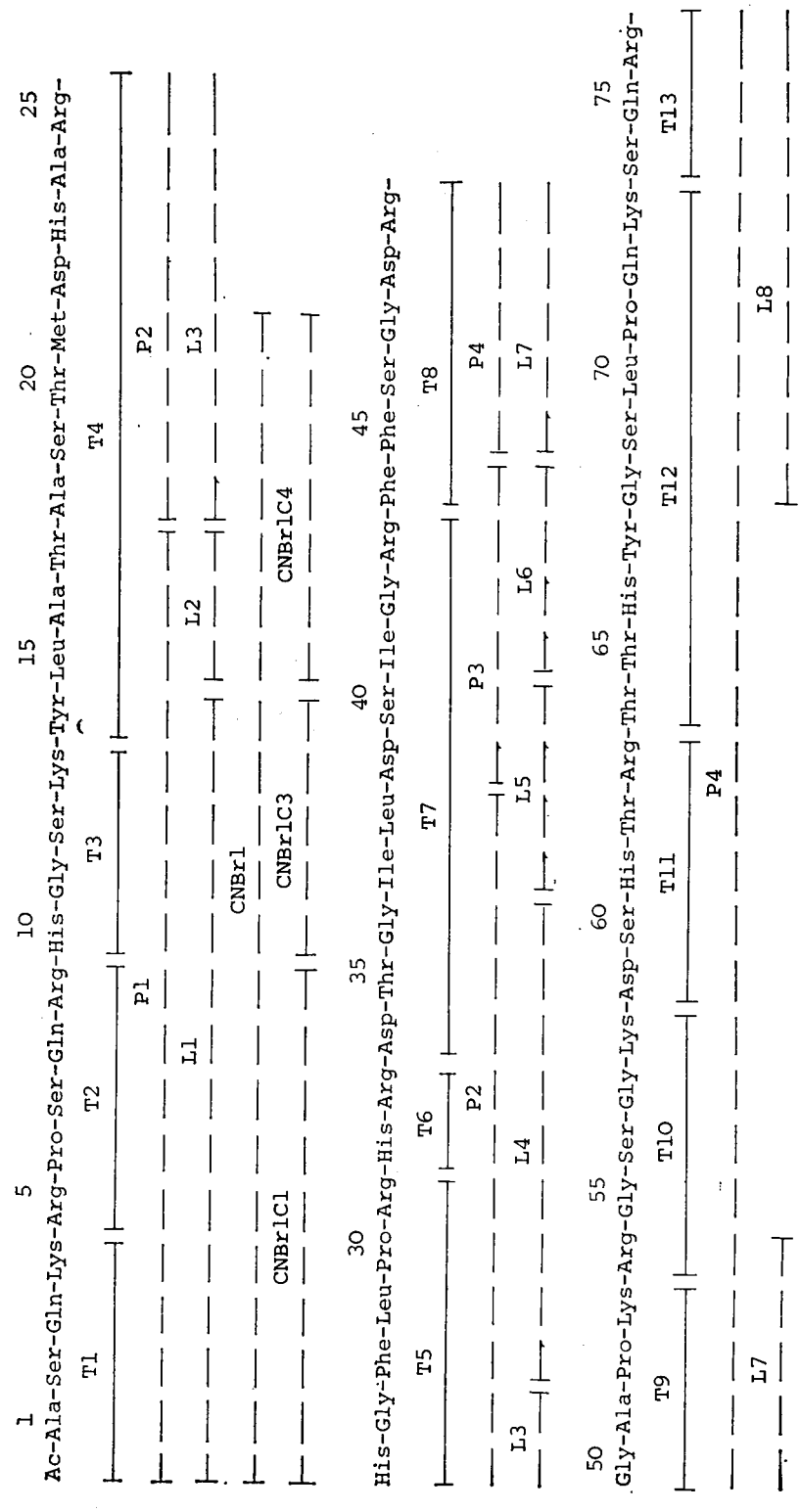
Fig. 1. *Tryptic peptides from the smaller basic protein of rat myelin*

——, Composition determined quantitatively; ——›, residue identified by dansyl-Edman degradation; —*›, residue identified by subtractive Edman method; ‹——, residue identified by digestion with carboxypeptidases A and B. Amide residues were assigned by electrophoretic mobility and were supported by enzymic digestion when required. The Supplementary Publication SUP 50029 contains full details (amino acid analyses, electrophoretic mobilities and/or chromatographic properties, and assignment of amide residues) for each tryptic peptide depicted in Fig. 1. Peptides derived from the tryptic peptides by enzymic digestion are also documented. Evidence for the sequence of the more complex peptides is discussed in the text.

residues (Carnegie 1971*a*; Supplementary Publication SUP 50029). Isolation of a basic peptide T14L1A1 (*m* 0.15) from an aminopeptidase O digestion of peptide T14L1 confirmed the assignment of asparagine at residue 5 in the latter peptide. This contrasts with the human protein, where the sequence Gln-Asp was found at residues 4 and 5 in the corresponding peptide (Carnegie, 1971*a*). The amino acid composition and sequence of peptides T14L1, T14L2, T14L1A1 (Fig. 1) and peptides L8 and L9 (Fig. 2) established the amino acid sequence in this region of the protein.

After one Edman degradation of peptide T15 the dansyl reaction produced Dns-Ile-Val as well as Dns-Ile, owing to the slow rate of hydrolysis of the Ile-Val bond.

A neutral peptide that gave the same colour with cadmium–ninhydrin and had an amino acid analysis identical with basic peptide T16, was isolated in


1 5 10 15 20 25
Ac-Ala-Ser-Gln-Lys-Arg-Pro-Ser-Gln-Arg-His-Gly-Ser-Lys-Tyr-Leu-Ala-Thr-Ala-Ser-Thr-Met-Asp-His-Ala-Arg-
T1
T2
T3
T4
P1
P2
L1
L2
L3
CNBr1
CNBr1C1
CNBr1C3
CNBr1C4
30 35 40 45
His-Gly-Phe-Leu-Pro-Arg-His-Arg-Asp-Thr-Gly-Ile-Leu-Asp-Ser-Ile-Gly-Arg-Phe-Phe-Ser-Gly-Asp-Arg-
T5
T6
T7
T8
P2
P3
P4
L3
L4
L5
L6
L7
50 55 60 65 70 75
Gly-Ala-Pro-Lys-Arg-Gly-Ser-Gly-Lys-Asp-Ser-His-Thr-Arg-Thr-Thr-His-Tyr-Gly-Ser-Leu-Pro-Gln-Lys-Ser-Gln-Arg-
T9
T10
T11
T12
T13
P4
L7
L8

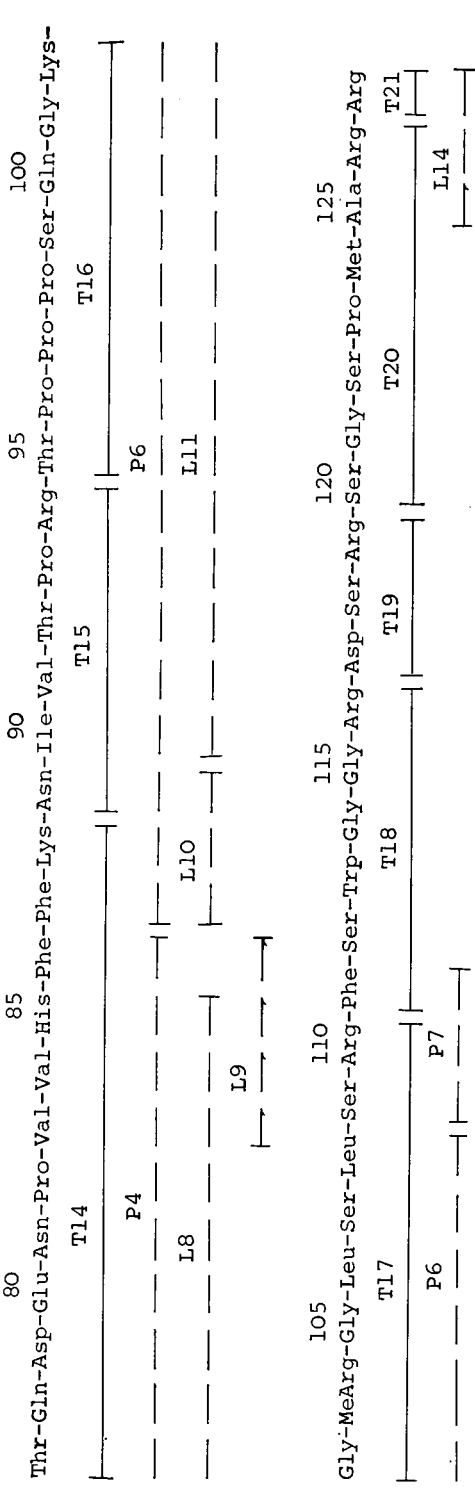


Fig. 2. *Complete amino acid sequence of the smaller basic protein from rat brain myelin*

Tryptic (T), peptic (P), thermolytic (L) and cyanogen bromide (CNBr) peptides are indicated. ——, Composition determined quantitatively; ——, residue identified by dansyl-Edman method. Full details (amino acid analyses, electrophoretic mobilities, chromatographic properties) for each peptide depicted in Fig. 2, are presented in the Supplementary Publication SUP 50029.

```
                    1                   10                  20                  30                  40
Rat S protein       Ac A S - Q K R P S Q R H G S K Y L A T A S T M D H A R H G F L P R H R D T G I L D
Rat L protein
Rabbit              Ac(A.S.-.Q)K/R.P.S.Q)R/H.G.S)K/Y.L.A.T.A.S.T.M.D.H.A)R/H.G.F.L.P)R/H R/D.T.G.I.L.D
Guinea pig
Pig                   Q A S K(R.P.S.Q)R.H G S K.Y L A S(A.S.T)M B H A R H G F L P R H R D T G I L D
Bovine              Ac A S A Q K R P S Q R - - S K Y L A S A S T M D H A R H G F L P R H R D T G I L D
Monkey
Human               Ac A S - Q K R P S Q R H G S K Y L A T A S T M D H A R H G F L P R H R D T G I L D
Invariant           ├──   ──────────    ──────  ──────────────────────────────────────────────────•

                   41                  50                  60                  70                  80
Rat S protein       S I G R F F S G D R G A P K R G S G K D S H - - T R T T H Y G S L P Q K S Q - -
Rat L protein
Rabbit              S.I.G)R/F F S S D R G A P K R G S G K D - - H A A R T T H Y G S L P Q K S - G H
Guinea pig                /F G S D R A A P K R G S.G K D S H H A A R T T H Y G S L P Q K S Q - -
Pig                 S L G R F F G A D R.G A P K.R G S G K/B -,H,-,A,-,R.T T(H.Y.G.S.L.P.Q)K/A.Z.G.H
Bovine              S L G R F F G S D R G A P K R G S G K D G H H A A R T T H Y G S L P Q K A Q G H
Monkey                    /F G G D R G V P K R G S G K D S H H A A R T A H Y G S L P Q K S - G H
Human               S I G R F F G G D R G A P K R G S G K D S H H P A R T A H Y G S L P Q K S H G -
Invariant           ─ ──────  ────  ──────────────────  ──  ──────────────────

                   81                  90                  100                 110                 120
Rat S protein       R T Q D E N P V V H F F K N I V T P R T P P P S Q G K G R G L S L S R F S W G -
Rat L protein                                                                                 /F.S.W.G.A.
Rabbit              R P Q D E N P V V H F.F)K/N.I.V.T.P)R/T.P.P.P.S.Q.G)K/G.R.G.L.S.V.T)R/F.S.W.G.A.
Guinea pig          R S Q D E N P V V H/
Pig                 R.P.Z.B.Z.B.P.V.G.H.F.F)K/N I V T P R/T.P.P.-.S.Z.G)K G R G L S L S R/
Bovine              R P Q D E N P V V H F F K N I V T P R T P P P S Q G K G R G L S L S R F S W G A
Monkey              R T Q D E N P V V H/
Human               R T Q D Q D P V V H F F K N I V T P R T P P P S Q G K G R G L S L S R F S W G A
Invariant           ─ ──  ──  ──  ──────────────────────  ──────────────────  ──────────

                   121                 130                 140                 150                 160
Rat S protein       - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Rat L protein       E.G.Q)K/                  /A.S.D.Y)K/S.A.H)K/G.F)K/G.A.-.D.A.Q.G.T.L.S)K/I.F)K/L.G.
Rabbit              E.G.Q)K/P.G.F.G.Y.G.G)R/A.A.D.Y)K/S.A.H)K/G.L)K/G.A.-.D.A.Q.G.T.L.S)R/L.F)K/L.G.
Guinea pig
Pig                               /A.P.B.Y.K.P.A.H)K/G L K/G A Q D A G Q T L - K/I F K/L G
Bovine              E G Q K P G F G Y G G R A S D Y K S A H K G L K G H - D A Q G T L S K I F K L G
Monkey
Human               E G Q R P G F G Y G G R A S D Y K S A H K G F K G V - D A Q G T L S K I F K L G
Invariant
```

	161 170
Rat S protein	G R D S R S G S P M A R R
Rat L protein	G)R/D.S)R/
Rabbit	G)R/D.S)R/S.G.S.P.M.A)R/R
Guinea pig	
Pig	G R/D S R/- G(S.P.M.A)R/R
Bovine	G R D S R S G S P M A R R
Monkey	
Human	G R D S R S G S P M A R R
Invariant	——— ———

Fig. 3. *Comparison of amino acid sequences of myelin basic proteins from different species*

The single letter code used for the amino acids is as follows: A, alanine; B, aspartic acid or asparagine, not distinguished; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine, K, lysine; L, Leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; X, undetermined or atypical amino acid; Y, tyrosine; Z, either glutamic acid or glutamine, not distinguished. The punctuation is that used by Dayhoff (1972). Invariant regions are shown ——; however, it should be noted that the data are incomplete for some species. The amino acid-sequence data presented in this figure were obtained from a number of reports: the rat L protein (Martenson *et al.*, 1972*b*), the rabbit protein (Brostoff & Eylar, 1972), the guinea pig, rabbit and monkey proteins (Shapira *et al.*, 1970), the pig protein (Kornguth *et al.*, 1972, Martenson *et al.*, 1972*a*), the bovine protein (Eylar *et al.*, 1971) and the human protein (Carnegie, 1971*a*).

approximately one-fifth the yield of peptide T16. The mobility of this peptide was consistent with it having been derived from peptide T16 by deamidation of glutamine (Offord, 1966).

Peptide T17 was the only tryptic peptide in the S protein which contained methylarginine. The related peptides T17a and T17b (Fig. 1) were also isolated and they are discussed below. Digestion of peptides T17 and T17b with thermolysin yielded three peptides T17L1, T17bL1 and T17bL2 whose amino acid composition and sequence established the sequence of the parent peptide. A peptide L12 (Fig. 2) with mobility at pH6.5 and amino acid composition equivalent to peptide T17bL2 was found in a thermolysin digest of the whole S protein.

Peptide T18 was the only tryptic peptide in the S protein which contained tryptophan, as judged by the Ehrlich procedure. Aminopeptidase O digestion of this peptide yielded only phenylalanine and serine residues in 1.0:1.0 ratio. Phenylalanine was shown to be the *N*-terminal amino acid by the dansyl technique. The Edman-degradation products of peptide T18 after three (3T18) and four (4T18) steps were characterized by their amino acid composition, mobility at pH6.5 and *N*-terminal amino acid. Peptide 3T18 (*m* 0.63) had an *N*-terminal glycine residue and an amino acid composition glycine 1.8, arginine 1.0; its mobility was consistent with the absence of tryptophan. Peptide 4T18 (*m* 0.69) had an *N*-terminal glycine residue and the amino acid composition glycine 1.2, arginine 1.0. These results are consistent with the sequence of peptide T18 shown in Fig. 1.

Peptide T20 contained methionine but was isolated in only one form, not two, as was peptide T4.

Free arginine (T21) was found in yields greater than would be expected from the slow hydrolysis of peptide T10.

Overlap of tryptic peptides

Digestion of the rat S protein with CNBr, pepsin and thermolysin was used to obtain peptides that would overlap the tryptic peptides. Numerous peptides were isolated, analysed and their *N*-terminal residues determined. Only a limited number were selected as evidence for the overlapping of the tryptic peptides based on the uniqueness of the less common residues. In certain regions additional peptides were used as further confirmation of the alignment. The peptic (P) and thermolytic (L) overlap peptides are numbered from the *N*-terminal end for convenience, and they are summarized in Fig. 2. Full details of the data and arguments used to assign the amino acid sequence of the S protein are presented in the Supplementary Publication SUP 50029.

In none of the tryptic, peptic or thermolytic digests of the rat S protein were any peptides found that would have corresponded to peptides from the sequence of amino acids between residues 120 and 160 in the other myelin basic proteins (Fig. 3). The

proposed sequence of the rat S protein is shown in Fig. 2.

Methylation of arginine-104 in the rat S protein

As found in the human (Baldwin & Carnegie, 1971*a*) and bovine (Brostoff & Eylar, 1971) basic proteins the rat S protein contained a single arginine residue (104) which was methylated. Peptide T17 contained ω-*N*-monoethylarginine and ω-*NN'*-dimethylarginine. No trace of ω-*NN*-dimethylarginine was found. Trypsin does not appear to hydrolyse peptide bonds at methylated arginine residues (Baldwin & Carnegie, 1971*b*; Sundarraj & Pfeiffer, 1973) and isolation of peptide T17a and T17b in yields comparable with that of peptide T17 indicated that arginine-104 was not completely methylated. No trace of a Gly-MeArg peptide was found, further indicating the inactivity of trypsin at methylarginine residues.

Theoretically it should be possible to calculate the content of methylarginine in the S protein from the yields of peptides T17a (Gly-Arg) compared with that of peptide T17. However, the steps required to obtain peptides T17 and T17a in homogeneous state are not identical and the yields thus give only a rough estimate of the amounts of methylarginine. It was necessary to obtain the methylarginine content from a single peptide which contained both the methylated and unmethylated forms of residue 104. Peptide L11 (Fig. 2) was such a peptide because thermolysin does not hydrolyse peptide bonds at arginine residues. The proportion of arginine to ω-*N*-monomethylarginine to ω-*NN'*-dimethylarginine at residue 104 was calculated to be 4:4:1. Results obtained with peptide P6 (Fig. 2) were consistent with this ratio.

Discussion

Amino acid sequence of the rat S protein

In Fig. 3 the amino acid sequence of the rat S protein is compared with the complete, or partial, sequences of basic proteins from six other species. To achieve maximum homology between sequences gaps were introduced and regions where no evidence has been reported are left blank (Dayhoff, 1972). When an amino acid is referred to in the discussion, the number given indicates the position of that amino acid in the sequence shown in Fig. 3.

Part of the amino acid sequence of the S protein has been reported in connexion with studies on localization of the determinant which induces experimental autoimmune encephalomyelitis in rats (Dunkley *et al.*, 1973; McFarlin *et al.*, 1973). There is disagreement between the sequence found by McFarlin *et al.* (1973) and that presented in Fig. 3 at four residues. The glycine at position 85 and alanine at position 51 reported by McFarlin *et al.* (1973) are printing errors and should read, as in Fig. 3, glutamic acid and glycine respectively (R. F. Kibler, personal communication). In the present study assignment of -Ser-Gly- at positions 47 and 48, rather than -Gly-Ser (McFarlin *et al.*, 1973), was based on both dansyl–Edman and subtractive Edman techniques.

(1) *The major deletion.* The rat S protein differs from all other myelin basic proteins in having a major internal deletion of 40 amino acid residues equivalent to residues 120–160 in the larger proteins (Fig. 3). Whether it is glycine-119 or glycine-160 that is deleted cannot be resolved. This large deletion is unlike those normally found in other families of proteins because of its size and internal location. The central position of the deletion ensures that the S protein is not derived from the L protein by proteolysis. As both rat proteins are normally expressed in inbred and outbred rat strains they must be products of non-allelic genes (Martenson *et al.*, 1971*b*).

How could the S-protein gene have been formed? A number of related but different haemoglobin chains have been formed by gene duplication followed by separate evolution of the individual genes, but in each case the products have retained approximately the same length (Dayhoff, 1972). Although the S-protein gene was probably produced by a duplication process the shorter length of the product necessitates that there must have been excision of a portion of the gene at the time of duplication or at some later stage. Two mechanisms could explain the formation of the smaller S-protein gene. The first involves unequal crossover at partially homologous DNA regions during synapsis. Bauer (1972) claimed that many regions of the basic protein, including those coding for amino acid residues 114–124 and 154–164, have homology with an 'ancestral histone IV gene'. An error in alignment of these regions at synapsis could lead to a new gene coding for a protein equivalent in size and sequence to the S protein. The second possible mechanism for formation of the S-protein gene is by misrepair of DNA coding for the rat L protein after breakage by either enzymic or mechanical action. Such a mechanism has been envisaged for the formation of the immunoglobulin Sac (Smithies *et al.*, 1971).

(2) *Comparison with other myelin basic proteins.* The close homology in sequence between the rat S protein and other myelin basic proteins is shown in Fig. 3. A large proportion of the basic protein sequence appears to be invariant. Dayhoff (1972) calculated the mutation acceptance rate and showed that the rate was low for myelin basic proteins when compared with haemoglobins and growth hormones but higher than that for cytochromes *c*. The regions of conserved sequence are likely to be functionally or structurally important. In this respect it is noteworthy that the binding of cerebroside sulphatide to bovine

basic protein found *in vitro* (London *et al.*, 1973) does not appear to correlate with conserved regions, as would be expected if this binding was essential for myelin structure.

A conservative region in the protein (residues 90–120) contains a number of interesting residues: a threonine residue (100) that can be glycosylated *in vitro* (Hagopian *et al.*, 1971), an unusual triproline sequence (101–103), a serine residue (112) that can be phosphorylated *in vitro* (Carnegie *et al.*, 1973), a methylated arginine residue (109), and a tryptophan residue (118) that is essential for the induction of experimental autoimmune encephalomyelitis in guinea pigs and which interacts with serotonin *in vitro* (Carnegie *et al.*, 1972).

In the regions of sequence variability the amino acid replacements are typical of those found in other proteins and require only single base changes to be expressed (Dayhoff, 1972). The only exceptions are at residues 143 and 146. With other proteins deletions are frequently found at the *N*- and *C*-terminal ends (Dayhoff, 1972), but the relatively conservative sequence at the *N*- and *C*-terminal ends of the myelin basic proteins suggests that these regions may be of functional importance.

(3) *Relationship of rat S to rat L protein.* A number of studies have shown that in adult rats the L protein is quantitatively less important than the S protein. The ratio varied depending on whether the proteins were extracted from whole brain (Martenson *et al.*, 1970*a*; Eng *et al.*, 1971) or isolated myelin (Sammeck *et al.*, 1971; Adams & Osborne, 1973) but was always between 1:1.75 and 1:3. The ratio did not vary in preparations from different regions of the adult central nervous system (Sammeck *et al.*, 1971). In developing brain there was proportionally more L protein than S protein (Adams & Osborne, 1973) but the total amount of basic protein maintained a constant ratio to the proteolipid protein during development of myelin. Pfeiffer & Wechsler (1972) have isolated a neoplastic clone of Schwann cells which, although of peripheral origin, produced a single basic protein closely similar in amino acid composition and electrophoretic properties to the L protein. Thus it is possible that in the normal central nervous system the L and S proteins are produced by different cells.

Little information is available on the amino acid sequence of the L protein, apart from preliminary studies on tryptic peptides which indicated that there is reasonable homology with the human and bovine basic proteins (Martenson *et al.*, 1972*b*; P. R. Dunkley, unpublished work). Possibly the 40 residues that are deleted in the S protein are not required for whatever structural or functional role the basic protein plays in myelin, but if this region were unimportant it would be expected to be variable in sequence, and this is not the case (Fig. 3). Perhaps there is an advantage in retaining two basic proteins with common sequence as mutational errors leading to loss of function would be minimized.

Heterogeneity of the rat S protein

Myelin basic proteins, including both the S and L proteins, that appear homogeneous on polyacrylamide-gel electrophoresis at acid pH values show heterogeneity at alkaline pH values (Martenson *et al.*, 1970*b*). Five bands of protein are usually seen and these have been isolated (Martenson *et al.*, 1971*c*). The basis of this heterogeneity is unknown, although a number of suggestions have been made; these include methylation of an arginine residue (Carnegie, 1971*b*), deamidation of glutamine or asparagine residue(s) (Martenson *et al.*, 1971*c*) and limited proteolysis at the *C*-terminal end of the protein (Bergstrand, 1971).

A single arginine residue in basic proteins from a number of species is known to be methylated. Previous studies on the methylarginine content of rat basic proteins were on mixtures of L and S proteins (Brostoff & Eylar, 1971; Deibler & Martenson, 1973*a*). In the present study the S protein was found to be a mixture of three proteins differing in the extent of methylation of arginine-109. The proportions of proteins containing arginine, ω-*N*-monomethylarginine and ω-*NN'*-dimethylarginine at this position were shown to be 4:4:1. The rat has a lower proportion of ω-*NN'*-dimethylarginine than the human protein (Baldwin & Carnegie, 1971*b*). Since the ratio of ω-*N*-monomethylarginine to ω-*NN'*-dimethylarginine does not appear to vary from batch to batch nor from laboratory to laboratory for any given species (P. R. Carnegie & P. R. Dunkley, unpublished work), this would suggest that the extent of methylation must be genetically regulated within the species by some unknown mechanism. However, partial methylation of arginine does not account for the heterogeneity of the proteins at alkaline pH values (Deibler & Martenson, 1973*b*).

In the present study a number of anomalous tryptic peptides were obtained in low yield (see the Results section). One of these was derived from peptide T4 by partial oxidation of methionine, which is a well-known artifact of peptide isolation (Harris & Roos, 1959).

A second peptide had the same composition as peptide T16, but was neutral rather than basic, a change that would suggest deamidation. This was the only peptide found with this type of change, and it is noteworthy that a similar yellow spot was observed on peptide 'maps' of human, bovine and rat L basic proteins (P. R. Dunkley, unpublished work). This deamidation may contribute to the heterogeneity of the basic proteins displayed at alkaline pH.

A third peptide had the same composition as peptide T11 (Fig. 1) but was acidic rather than basic. It is possible that this change in mobility was due to phosphorylation. When S protein was used as a substrate for a protein kinase from rabbit muscle three major sites were phosphorylated (residues 36, 57 and 112) but the region containing peptide T11 was not one of them (Carnegie *et al.*, 1973; Carnegie *et al.*, 1974). However, rat basic proteins can be phosphorylated *in vivo* but the sites of phosphorylation have yet to be determined (Miyamoto & Kakiuchi, 1974). A change in electrophoretic pattern, from homogeneous at acid pH to heterogeneous at alkaline pH values, was shown for ovalbumin to be due to phosphorylation (Perlmann, 1952).

The bovine basic protein may exist in two forms, one with histidine at position 24 and one with leucine at this position (Kornguth *et al.*, 1972). No peptides from the S protein were found in sufficient yield to indicate that this type of heterogeneity exists, nor was there any evidence for or against the possibility of limited proteolysis at the *C*-terminal end of the basic protein as was suggested by Bergstrand (1971). Perhaps the heterogeneity observed by Martenson *et al.* (1970*b*) is not due to any one type of change but rather to a combination of the types of heterogeneity discussed above.

Although the amino acid sequence of the myelin basic protein is well established the role of this protein in myelin structure is unknown. However, the marked clinical symptoms which appear in experimental autoimmune encephalomyelitis would suggest that this protein has a vital function in the central nervous system.

We acknowledge financial support from the National Multiple Sclerosis Society, New York, U.S.A. (grant no. 719-A-5) and the Australian Research Grants Committee. Dr. J. Thompson is thanked for helpful discussions and Mr. G. Luxford and Mr. D. Mudie are thanked for technical assistance.

References

Adams, D. H. & Osborne, J. (1973) *Neurobiology* **3**, 91–112

Alvord, E. C. (1968) in *The Central Nervous System* (Bailey, O. T. R. & Smith, D. E., eds.), pp. 52–70, Williams and Wilkins Co., Baltimore

Baldwin, G. S. & Carnegie, P. R. (1971*a*) *Science* **171**, 579–581

Baldwin, G. S. & Carnegie, P. R. (1971*b*) *Biochem. J.* **123**, 69–74

Bauer, K. (1972) *Biochem. J.* **126**, 1245–1248

Bergstrand, H. (1971) *Eur. J. Biochem.* **21**, 116–124

Brostoff, S. & Eylar, E. H. (1971) *Proc. Nat. Acad. Sci. U.S.* **68**, 765–769

Brostoff, S. & Eylar, E. H. (1972) *Arch. Biochem. Biophys.* **153**, 590–598

Carnegie, P. R. (1969) *Biochem. J.* **111**, 240–242

Carnegie, P. R. (1971*a*) *Biochem. J.* **123**, 57–67

Carnegie, P. R. (1971*b*) *Nature (London)* **229**, 25–28

Carnegie, P. R. & Dunkley, P. R. (1973) *Abstr. Meet. Int. Soc. Neurochem. 4th*, pp. 67–68

Carnegie, P. R., Lamoureux, G. & Bencina, B. (1967) *Nature (London)* **214**, 407–408

Carnegie, P. R., Smithies, J. R., Caspary, E. A. & Field, E. J. (1972) *Nature (London)* **240**, 561–563

Carnegie, P. R., Kemp, B. E., Dunkley, P. R. & Murray, A. W. (1973) *Biochem. J.* **135**, 569–572

Carnegie, P. R., Dunkley, P. R., Kemp, B. E. & Murray, A. W. (1974) *Nature (London)* in the press

Cotman, C. W. & Mahler, H. R. (1967) *Arch. Biochem. Biophys.* **120**, 384–396

Dayhoff, M. D. (1972) in *Atlas of Protein Sequence and Structure*, vol. 5, National Biomedical Research Foundation, Silver Spring, Md.

Deibler, G. E. & Martenson, R. E. (1973*a*) *J. Biol. Chem.* **248**, 2387–2391

Deibler, G. E. & Martenson, R. E. (1973*b*) *J. Biol. Chem.* **248**, 2392–2396

Deibler, G. E., Martenson, R. E. & Kies, M. W. (1970) *Biochim. Biophys. Acta* **200**, 342–352

Deibler, G. E., Martenson, R. E. & Kies, M. W. (1972) *Prep. Biochem.* **2**, 139–165

Dunkley, P. R. & Carnegie, P. R. (1974) in *Research Methods in Neurochemistry* (Marks, N. & Rodnight, R., eds.), vol. 2, pp. 219–245, Plenum Press, New York

Dunkley, P. R., Coates, A. S. & Carnegie, P. R. (1972) *Proc. Aust. Biochem. Soc.* **5**, 37

Dunkley, P. R., Coates, A. S. & Carnegie, P. R. (1973) *J. Immunol.* **110**, 1699–1701

Easley, C. W. (1965) *Biochim. Biophys. Acta* **107**, 386–388

Eng, L. F., Chao, F.-C., Gerstl, B., Pratt, D. & Tavaststjerna, M. G. (1968) *Biochemistry* **7**, 4455–4465

Eng, L. F., Bond, P. & Gerstl, B. (1971) *Neurobiology* **1**, 58–63

Eylar, E. H., Brostoff, S., Hashim, G., Caccam, J. & Burnett, P. (1971) *J. Biol. Chem.* **246**, 5770–5784

Gaitonde, M. K. & Dovey, T. (1970) *Biochem. J.* **117**, 907–911

Hagopian, A., Westall, F. C., Whitehead, J. S. & Eylar, E. H. (1971) *J. Biol. Chem.* **246**, 2519–2523

Harris, J. I. & Roos, P. (1959) *Biochem. J.* **71**, 434–445

Hartley, B. S. (1970) *Biochem. J.* **119**, 805–822

Hill, R. L. (1965) *Advan. Protein Chem.* **20**, 37–107

Hirs, C. H. W. (1956) *J. Biol. Chem.* **219**, 611–621

Kornguth, S. E., Kozel, L. R. & Smithies, O. (1972) *Nature (London)* **237**, 49–50

London, Y., Demel, R. A., Geurtz von Kessell, W. S. M., Vossenberg, F. G. A. & Van Deenan, L. L. M. (1973) *Biochim. Biophys. Acta* **311**, 520–530

Martenson, R. E., Deibler, G. E. & Kies, M. W. (1970*a*) *Biochim. Biophys. Acta* **200**, 353–362

Martenson, R. E., Deibler, G. E. & Kies, M. W. (1970*b*) *J. Neurochem.* **17**, 1329–1330

Martenson, R. E., Deibler, G. E. & Kies, M. W. (1971*a*) *J. Neurochem.* **18**, 2417–2426

Martenson, R. E., Deibler, G. E. & Kies, M. W. (1971*b*) *J. Neurochem.* **18**, 2427–2433

Martenson, R. E., Deibler, G. E. & Kies, M. W. (1971*c*) in *Immunological Disorders of the Nervous System*, vol. 49, pp. 76–94, Research Publication, Association for Research in Nervous and Mental Diseases

Martenson, R. E., Deibler, G. E. & Kies, M. W. (1972*a*) *Nature (London)* **234**, 87–89
Martenson, R. E., Deibler, G. E., Kies, M. W., McKneally, S. S., Shapira, R. & Kibler, R. F. (1972*b*) *Biochim. Biophys. Acta* **263**, 193–203
McFarlin, D. E., Blank, S. E., Kibler, R. F., McKneally, S. & Shapira, R. (1973) *Science* **179**, 478–479
Mehl, E. (1967) *Abstr. Congr. Neurochem. 1st*, p. 154
Miyamoto, E. & Kakiuchi, S. (1974) *J. Biol. Chem.* in the press
Offord, R. E. (1966) *Nature (London)* **211**, 591–593
Perlman, G. E. (1952) *J. Gen. Physiol.* **35**, 711–726
Pfeiffer, S. E. & Wechsler, W. (1972) *Proc. Nat. Acad. Sci. U.S.* **69**, 2885–2889
Sammeck, R., Martenson, R. E. & Brady, R. O. (1971) *Brain Res.* **34**, 241–245
Shapira, R., McKneally, S. S., Chou, F.-C. & Kibler, R. F. (1970) *J. Biol. Chem.* **246**, 4630–4640
Shotton, D. M. & Hartley, B. S. (1973) *Biochem. J.* **131**, 643–675
Smithies, O., Gibson, D. M., Fanning, E. M., Percy, M. E., Parr, D. M. & Connell, G. E. (1971) *Science* **172**, 574–577
Sundarraj, N. & Pfeiffer, S. E. (1973) *Biochem. Biophys. Res. Commun.* **52**, 1039–1045