Biochem. J. (1979) **181**, 659–665
Printed in Great Britain

The Isolation and Partial Sequence of Peptides Produced by Cyanogen Bromide Cleavage of Calf Thymus Non-Histone Chromosomal High-Mobility-Group Protein 2

SEQUENCE HOMOLOGY WITH NON-HISTONE CHROMOSOMAL HIGH-MOBILITY-GROUP PROTEIN 1

By John M. WALKER, Keith GOODERHAM and Ernest W. JOHNS
Chester Beatty Research Institute of Cancer Research, Royal Cancer Hospital, Fulham Road, London SW3 6JB, U.K.

(*Received* 22 *March* 1979)

Peptides produced by CNBr cleavage of non-histone chromosomal protein HMG 2 (CNBr peptides) were isolated and characterized, and their partial sequences were determined. The present sequence data account for over half of the sequence of the protein HMG (high-mobility-group) 2 molecule, and, together with previously published results, provide interesting information on the charge distribution within the molecule. Comparison of the CNBr-peptide-sequence data for protein HMG 2 with the previously published data on the CNBr peptides from protein HMG 1 reveals extensive sequence homology between the two proteins. Detailed evidence for the amino acid-sequence data has been deposited as Supplementary Publication SUP 50095 (6 pages) at the British Library Lending Division, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K., from whom copies may be obtained on the terms given in *Biochem. J.* (1978) **169**, 5.

Recent models of the structure of chromatin suggest that it is composed of repeating subunits, each subunit or nucleosome consisting of about 200 base pairs of DNA associated with the five histones (Kornberg, 1974). It is thought that 140 base pairs of DNA are associated with an octamer of two each of the four histones H2A, H2B, H3 and H4 (Shaw *et al.*, 1976), whereas histone H1 is probably bound to the section of DNA linking the subunits (Varshavsky *et al.*, 1976). The amino acid sequences of four of the five histones have now been known for a number of years. However, until recently very little sequence information has been available for any of the non-histone chromosomal proteins involved in the nucleosome structure. During the past few years our laboratory has been studying a particular group of non-histone chromosomal proteins called the HMG proteins (Johns *et al.*, 1975). The presence of HMG proteins in a variety of organisms and tissues, including avian erythrocytes (Rabbani *et al.*, 1978; Sterner *et al.*, 1978), trout testis (Watson *et al.*, 1977), wheat and yeast (Spiker *et al.*, 1978) implies a widespread occurrence in eukaryotic nuclei. There are four main HMG proteins in thymus, HMG 1, 2, 14 and 17. All four of these proteins have been isolated in a pure form from both pig and calf thymus (Goodwin *et al.*, 1975, 1977*b*; Sanders, 1975), and have all been shown to be present in isolated nucleosomes (Goodwin *et al.*, 1977*a*). Because of the quantities of the HMG proteins present in the nucleus (10^5–10^6 molecules of each protein), we consider that the HMG proteins are structural proteins, possibly involved in the higher order structure of the chromatin, and not involved in specific gene control.

We have previously published the complete amino acid sequences of calf thymus HMG proteins 14 and 17 (Walker *et al.*, 1977, 1979). As a further part of our characterization of the HMG proteins, we have determined the amino acid sequences of HMG proteins 1 and 2.

The molecular weights of HMG proteins 1 and 2, as determined by sedimentation equilibrium, are both approx. 26000 (Johns *et al.*, 1975). One of the noteworthy features of HMG proteins 1 and 2 is that over 50% of their amino acid residues are charged. Like the histones, 25% of the residues in both proteins are basic. However, unlike the histones, both proteins also contain 30% acidic amino acids (Johns *et al.*, 1975).

We report in the present paper the isolation and partial sequence of the peptides isolated from CNBr cleavage of protein HMG 2. The sequence data presented here account for over half the sequence of the HMG 2 protein molecule and provides interesting information on the charge distribution within the molecule. Comparison of the CNBr-peptide-sequence data for protein HMG 2 with the previously published

Abbreviations used: HMG protein, high-mobility-group protein; dansyl, 5-dimethylaminonaphthalene-1-sulphonyl.

data on the CNBr peptides from protein HMG 1 reveals extensive sequence homology between the two proteins. The significance of this homology is discussed.

Experimental

Materials

NN-Dimethyl-*N*-allylamine was Sequenator grade from Pierce Chemicals (Rockford, IL, U.S.A.). Benzene, n-butyl chloride and phenyl isothiocyanate for sequenator use were obtained from Rathburn Chemicals (Walkerburn Peeblesshire, Scotland). CNBr was obtained from Koch–Light (Colnbrook, Bucks., U.K.) Sephadex G-50 (Superfine grade) was obtained from Pharmacia (London W.5., U.K.). Carboxymethylcellulose (CM 52) ion-exchange resin and chromatography paper (no. 3MM) were obtained from Whatman (Maidstone, Kent, U.K.). Dansyl chloride was obtained as a 10% solution in acetone from Pierce Chemicals. Thin-layer polyamide plates were from BDH (Poole, Dorset, U.K.). Aluminium-backed thin-layer silica-gel 60 F_{254} plates were obtained from Merck. All other chemicals used were A.R. grade from BDH.

Methods

Isolation of protein HMG 2. Protein HMG 2 was prepared from calf thymus as described previously (Johns *et al.*, 1975).

CNBr cleavage of protein HMG 2 *and Sephadex-gel filtration of the cleavage products*

Protein HMG 2 (150mg) was dissolved in $(NH_4)_2CO_3$ solution (0.2M, 25ml) containing β-mercaptoethanol (0.7ml) and incubated at 37°C overnight. The solution was then freeze-dried and redissolved in 70% (v/v) formic acid (25ml). Solid CNBr (1g) was added with stirring, and the flask left under N_2 for 24h at room temperature. A further addition of CNBr (0.3g) was then made and the solutions left for a further 24h. After this time the protein solution was rotary-evaporated to dryness, redissolved in 0.01M-HCl (5ml) and passed through a column (2.4cm×100cm) of Sephadex G-50 (Superfine grade) previously equilibrated in 0.01M-HCl (containing 1ml of β-mercaptoethanol per litre). Fractions (8ml) were collected and the material under the peaks pooled as shown in Fig. 1. Pooled material was rotary-evaporated to dryness and stored at −20°C until required.

Ion-exchange fractionation of Sephadex peaks B and C

Material in Sephadex-gel-filtration peaks B and C (Fig. 1) were dissolved in 10ml of 0.01M-sodium acetate, pooled, and the pH adjusted to 3.4. The combined samples were loaded on to a column (1.6cm×12cm) of carboxymethylcellulose (CM 52) equilibrated in 0.01M-sodium acetate, pH3.4, at a flow rate of 1ml/min. Peptides were eluted by applying a linear gradient (2×700ml) of 0.05–0.35M-NaCl in the same buffer. Fractions (10ml) were collected and peaks pooled as shown in Fig. 2. Peptides were recovered from these pools by acidification to 0.03M-HCl, followed by precipitation with 6 vol. of acetone. Precipitated peptides were recovered by centrifugation, washed three times with acetone and finally dried under vacuum.

Polyacrylamide-gel electrophoresis

Polyacrylamide-gel electrophoresis was carried out as described previously (Goodwin *et al.*, 1975).

Amino acid analysis

Total amino acids were measured with a Rank–Hilger Chromaspek amino acid analyser. Samples

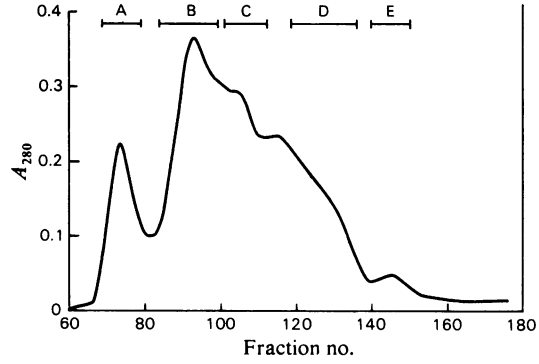

Fig. 1. *Sephadex G-50 elution profile of the CNBr-cleavage products from protein HMG* 2.

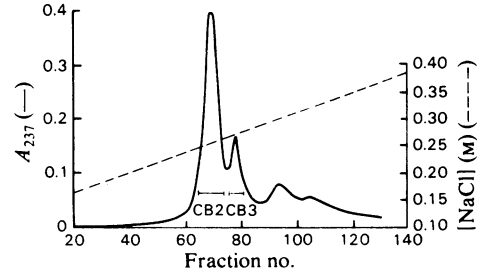

Fig. 2. *Ion-exchange chromatography of Sephadex G-50 pooled peaks B and C*

were hydrolysed in 6M-HCl at 110°C for 24h. No corrections were made for hydrolytic losses.

Small-peptide purification by high-voltage paper electrophoresis and paper chromatography

Small peptides (peptides CB5-1–CB5-4) were purified from gel-filtration fraction E by a combination of high-voltage paper electrophoresis at pH6.5 and descending paper chromatography in butanol/pyridine/acetic acid/water by using methods described previously (Walker *et al.*, 1977).

N-Terminal amino acid analysis and manual sequence determinations

Manual sequencing methods and identification of *N*-terminal amino acids by the dansyl method were carried out as described by Hartley (1970).

Automatic sequence determination

The sequence determination of large peptides (CB1–CB4) was carried out on a Beckman 890C sequencer, by using the standard *NN*-dimethyl-*N*-allylamine buffer program. The phenylthiohydantoin derivatives of the released amino acids were determined by both t.l.c. and back-hydrolysis to the free amino acids, with subsequent identification by amino acid analysis, as described previously (Walker *et al.*, 1978*b*). Peptide CB2 (8.0mg), peptide CB3 (4.4mg) and peptide CB4 (0.5mg) were subjected to sequenator analysis.

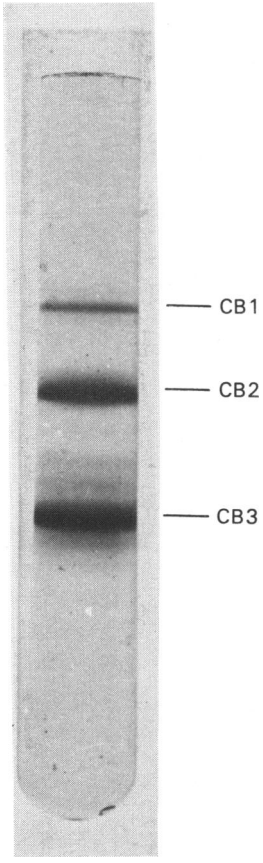

Fig. 3. 20% (*w*/*v*) *Polyacrylamide-gel-electrophoresis pattern of the CNBr-cleavage products of HMG* 2

Results

N-Terminal analysis of the CNBr-cleaved protein showed the presence of glycine, alanine, serine, lysine and proline. Polyacrylamide-gel electrophoresis of the total CNBr digest is shown in Fig. 3. Three main bands, CB1–CB3, can be seen. The first fraction (A) recovered from the Sephadex column run corresponds to the polyacrylamide-gel band CB1. The second and third fractions (B and C) comprised a mixture of peptides CB2 and CB3, which were resolved by ion-exchange chromatography (Fig. 2). Sephadex fraction D was further purified by descending paper chromatography for 3 days. Material remaining at the origin after this time was shown to have a single *N*-terminal amino acid and was designated 'peptide CB4'. Small amounts of other material present in fraction D that moved in this chromatography system were not amenable to further purification. Fraction E was resolved into four peptides by a combination of high-voltage paper electrophoresis (at pH6.5) and descending paper chromatography. These peptides are designated CB5-1–CB5-4 (Fig. 4). *N*-Terminal analysis showed peptides CB1 and CB3 to have a single *N*-terminal amino acid, lysine. No *N*-terminal amino acid was detected for peptide CB2, but this was later shown to be tryptophan (see peptide CB2 below). CB4 had serine as the *N*-terminal amino acid residue. These peptides represent the four major CNBr peptides isolated from protein HMG 2. The amino acid analysis of the four major peptides (CB1–CB4), and that of total protein HMG 2, are shown in Table 1. The main amino acids that characterize these peptides are underlined. The yields of these major peptides from cleavage of 150mg of protein HMG 1 were as follows: peptide CB1, 8mg; peptide CB2, 28mg; peptide CB3, 12mg; peptide CB4, 8mg. The sequence data obtained for all the CNBr peptides are presented in Fig. 4, together with the previously published *N*-terminal sequence of protein HMG 2. Also included in the Figure are the corresponding sequences from protein HMG 1 (Walker *et al.*, 1978*b*).

The results obtained for the individual CNBr peptides will be discussed separately.

Fig. 4. *The N-terminal sequences of protein HMG* 2, *peptides CB*2, *CB*3 *and CB*4 *obtained from protein HMG* 2, *and the corresponding protein-HMG* 1 *sequences*
Sequence differences between the two proteins are indicated by the 'boxed' regions.

Table 1. *Amino acid analysis of total protein HMG 2 and the pure CNBr fractions CB1, CB2, CB3 and CB4*
The main amino acids that characterize the protein or peptides are underlined.

Amino acid	Protein or peptide ...	Content (mol %)				
		HMG 2	CB1	CB2	CB3	CB4
Asp		9.9	10.2	11.6	8.5	7.6
Thr		2.0	1.7	1.8	2.4	4.8
Ser		6.6	5.7	5.3	6.6	13.8
Glu		18.7	17.7	29.8	8.9	17.5
Pro		7.5	11.6	5.8	13.2	4.4
Gly		6.1	9.6	7.0	8.5	2.8
Ala		7.5	9.8	9.2	6.3	6.0
Cys		+	0.4	—	1.0	0.8
Val		1.6	1.1	0.3	2.4	5.3
Met		1.9	—	—	—	—
Ile		1.4	2.3	1.6	2.9	—
Leu		1.8	3.4	1.8	4.6	—
Tyr		2.6	2.8	3.2	1.7	0.8
Phe		3.6	2.5	0.4	5.1	9.1
His		1.8	1.9	1.3	3.2	4.1
Lys		22.0	20.6	19.9	20.6	19.1
Arg		4.7	3.0	2.8	4.3	4.8
Homoserine		—	—	—	Present	Present
N-Terminal residue		Gly	Lys	Trp	Lys	Ser

Discussion

Peptide CB1

The amino acid analysis of this peptide is very similar to that of total protein HMG 2 (Table 1). Peptide CB1 would therefore appear to be produced by cleavage near to the *N*-terminus of protein HMG 2, and probably results from an absence of cleavage at other methionine residues within the molecule. Cleavage at a position considerably further down the protein chain would have resulted in a peptide with a much higher aspartic and glutamic acid content (see peptide CB2 below). The absence of homoserine or homoserine lactone in the amino acid analysis of this peptide confirms that the sequence of this peptide extends to the *C*-terminus of the molecule. Sequenator analysis of this peptide showed the *N*-terminal sequence to be the same as that for peptide CB3 (see below). Since greater information was obtained from the sequenator analysis of the smaller peptide, CB3, peptide CB1 provided no additional information, other than to place peptide CB3 in the *N*-terminal half of the molecule.

Peptide CB2

Peptide CB2 is characterized by a very high aspartic and glutamic acid content (41%) (Table 1), and a low content of hydrophobic and aromatic amino acids. A total of only five hydrophobic and aromatic amino acids are present in the 45-residue sequence of peptide CB2 (Fig. 4). This peptide is approx. 120 residues long (about half the size of the total HMG 2 molecule), and since the amino acid analysis shows an absence of homoserine or homoserine lactone, this peptide presumably derives from the *C*-terminus of the molecule. The amino acid sequence of the first 45 residues of this peptide (Fig. 4) shows that only ten aspartic and glutamic residues are located in this region. Since this peptide contains approx. 50 aspartic and glutamic residues, the majority of these residues must be present towards the *C*-terminus of this peptide. We have, in fact, previously shown for both HMG 2 and 1 proteins that there is a continuous sequence of approx. 40 aspartic and glutamic residues (HGA peptide) towards the end of this peptide (Walker *et al.*, 1978*a*). Comparison of peptide CB2 with the corresponding protein-HMG 1 sequence (Fig. 4) shows an amino acid insertion at position 27 in the protein-HMG 2 sequence. However, it was not possible to identify the amino acid residue at this position. Only a small increase in lysine was detected at this cycle, but an insufficient amount to identify lysine positively at this position. There is no obvious sequence homology between peptide CB2 and any of the histones. Recent results in our laboratory have shown that the HGA peptide starts seven residues beyond the sequence of peptide CB2 shown in Fig. 4.

Peptide CB3

In contrast with peptide CB2, peptide CB3 has a relatively low aspartic and glutamic acid content (17%), but a considerably higher content of hydrophobic and aromatic amino acids (17%; valine, isoleucine,

leucine, tyrosine and phenylalanine). The peptide is approx. 60 residues in length. The amino acid sequence of the first 45 residues is shown in Fig. 4. Difficulty was experienced in identifying certain residues towards the end of this sequence, but those that could be identified are included to allow comparison with the corresponding protein-HMG 1 sequence. The low aspartic and glutamic content of this peptide is reflected by the presence of only seven aspartic and glutamic residues in the first 45 residues. Six proline residues are present in the first 24 residues. Since proline residues necessarily produce a bend in a polypeptide chain, it seems unlikely that the region of the molecule that contains peptide CB3 will form an α-helical structure, particularly since other helix-breaking amino acids (glycine, serine and asparagine) are also present in this peptide. A strongly hydrophobic region with the sequence Phe-Phe-Leu-Phe-Ala is present between residues 26 and 31. The presence of a region enriched in hydrophobic amino acids is a common feature of the four histones (H2A, H2B, H3 and H4) involved in the nucleosome structure, where two each of the four histones are thought to bind together by hydrophobic interactions to form an octamer complex that is at the core of the nucleosome (Kornberg, 1974). It is therefore possible that this region of the protein-HMG 1 molecule is involved in hydrophobic interactions with other chromosomal proteins. No regions of sequence homology of greater than three residues can be seen with peptide CB3 and the known sequence of the histones.

Peptide CB4

At two attempts at sequenator analysis of peptide CB4, a considerable decrease in the yield of amino acid phenylthiohydantoin derivatives, after the valine residue at position 7, occurred. This was presumably due to partial cyclization of the newly liberated glutamine residue at position 8 to the corresponding pyroglutamyl derivative. Consequently, only limited information was obtained for this peptide. The peptide is obviously produced by cleavage at methionine at position 12 in the *N*-terminal sequence and provides an extension of the previously determined *N*-terminal sequence of protein HMG 2. The corresponding peptide was not isolated from the CNBr cleavage of protein HMG 1. However, a homologous peptide produced by pepsin cleavage of protein HMG 1 has been sequenced (J. M. Walker, unpublished work) and is included in Fig. 4 in order to allow further sequence comparison between HMG proteins 1 and 2.

Peptides CB5-1–CB5-4

Four small peptides were isolated from Sephadex-gel fraction E, and their amino acid sequences are included in Fig. 4 (peptides CB5-1–CB5-4). All four peptides were sequenced by the manual dansyl–Edman technique. Methionine residues were identified as a mixture of dansylhomoserine and dansylhomoserine lactone. All manual sequence data presented in Fig. 4 are consistent with the amino acid analyses of the purified peptides. Comparison of peptides CB5-3 and CB5-4 with the *N*-terminal sequence of protein HMG 2 (Fig. 4) shows these peptides to derive from the expected *N*-terminal CNBr peptide residues 1–12. They are obviously produced by cleavage at the aspartic acid–proline linkage at residues 4 and 5. The aspartic acid–proline linkage is known to be particularly susceptible to acid cleavage (Piszkiewicz *et al.*, 1970) and cleavage presumably occurred under the acid conditions of the CNBr-cleavage reaction. The yield of fraction E from the gel-filtration column (Fig. 1) was 6.0mg. The final yields of peptides from fraction E were all between 0.3 and 0.5mg.

The sequence data presented here, together with the previously determined *N*-terminal sequence and the known composition of the highly acidic region of peptide CB2, account for 75% of the total sequence of protein HMG 2.

In conclusion, therefore, a number of general features of the structure of protein HMG 2 may be deduced from the above sequence data.

Table 1 shows that all the large CB fractions (CB1–CB4) contain approx. 20% lysine, which is the same as the lysine composition in total protein HMG 2. Although short localized clusters of basic residues have been observed in the above sequence data (the most extreme example is the sequence Lys-Lys-Gly-Lys-Lys-Lys in peptide CB3), no extended regions of basic residues, similar to the highly grouped acidic residues found in peptide CB2, appear to occur in the sequence of protein HMG 2. It would therefore appear that, with the exception of the highly acidic region at the *C*-terminus of the molecule, the basic residues are fairly evenly distributed throughout the molecule. In contrast, the majority of the acidic residues are present as a continuous sequence of about 40 aspartic and glutamic residues towards the *C*-terminus of the molecule (peptide CB2). A function for this highly unusual acidic region has yet to be suggested. Two regions of relatively high concentration of hydrophobic amino acids also exist within the molecule (peptide CB3 and *N*-terminal sequence) and are both positioned in the *N*-terminal half of the molecule. Additionally, there would appear to be very little sequence homology between protein HMG 2 and the known sequences of the histones.

Sequence homology between HMG proteins 1 and 2

Initial analytical studies on proteins HMG 1 and HMG 2 suggested that the two proteins had similar

primary structures (Walker *et al.*, 1976). The sequence data presented in Fig. 4 reveal in detail the degree of sequence homology that exists between these two proteins. For the 135 amino acid residues that are directly comparable in this Figure, 110 amino acids are common to both proteins. Of the 25 differences that occur, only nine can be described as non-conservative changes. Particularly noteworthy is the conservation of sequence in two distinctly hydrophobic regions of the molecule: the sequence Tyr-Ala-Phe-Phe-Val present in the *N*-terminal sequence, and the sequence Phe-Phe-Leu-Phe-Ala present in peptide CB3. Additionally, we previously showed that both HMG 1 and 2 proteins have a continuous sequence of approx. 40 glutamic acid and aspartic acid residues (HGA peptide) in the *C*-terminal region of each molecule. Considering, therefore, the extensive sequence homology that we have demonstrated between the two proteins, and the fact that they both contain the unique structural feature of the HGA peptide, the question must arise as to whether or not proteins HMG 1 and 2 should be considered as two distinct proteins carrying out functionally separate roles within the structure of chromatin. It is possible that proteins HMG 1 and 2 might represent extreme forms of microheterogeneity in a single protein. Multiple forms of a chromosomal protein produced by microheterogeneity in the amino acid sequence has, of course, already been demonstrated in the case of histone H1 (Kinkade & Cole, 1966*a*,*b*), although in this case as well the reasons for the heterogeneity are not known. Unfortunately, the absence of a function for the HMG proteins other than that of 'structural' precludes the comparison of these two proteins in some form of biological assay. Until such time as this can be done, the reason for the presence in chromatin of two such closely related proteins must remain a mystery.

We thank Dr. J. R. B. Hastings for carrying out the amino acid analyses. Financial support for J. M. W. and E. W. J. from the Medical Research Council (grant no. 976/120) and for K. G. from the Institute of Cancer Research, is gratefully acknowledged.

References

Goodwin, G. H., Nicolas, R. H. & Johns, E. W. (1975) *Biochim. Biophys. Acta* **405**, 280–291
Goodwin, G. H., Woodhead, L. & Johns, E. W. (1977*a*) *FEBS Lett.* **73**, 85–88
Goodwin, G. H., Rabbani, A., Nicolas, R. H. & Johns, E. W. (1977*b*) *FEBS Lett.* **80**, 413–416
Hartley, B. S. (1970) *Biochem. J.* **119**, 805–815
Johns, E. W., Goodwin, G. H., Walker, J. M. & Sanders, C. (1975) *Ciba Found. Symp.* **28**, 95–108
Kinkade, J. M. & Cole, R. D. (1966*a*) *J. Biol. Chem.* **241**, 5790–5797
Kinkade, J. M. & Cole, R. D. (1966*b*) *J. Biol. Chem.* **241**, 5798–5805
Kornberg, R. D. (1974) *Science* **184**, 868–871
Piszkiewicz, D., Landon, M. & Smith, E. L. (1970) *Biochem. Biophys. Res. Commun.* **40**, 1173–1178
Rabbani, A., Goodwin, G. H. & Johns, E. W. (1978) *Biochem. Biophys. Res. Commun.* **81**, 351–358
Sanders, C. (1975) Ph.D. Thesis, University of London
Shaw, B. R., Herman, T. M., Kovacic, R. T., Beaudreau, G. S. & Van Holde, K. E. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 505–509
Spiker, S., Mardian, J. K. W. & Isenberg, I. (1978) *Biochem. Biophys. Res. Commun.* **82**, 129–135
Sterner, R., Boffa, L. C. & Vidali, G. (1978) *J. Biol. Chem.* **253**, 3830–3836
Varshavsky, A. J., Bakayev, V. V. & Georgiev, G. P. (1976) *Nucleic Acids Res.* **3**, 477–492
Watson, P. C., Peters, E. H. & Dixon, G. H. (1977) *Eur. J. Biochem.* **74**, 53–60
Walker, J. M., Goodwin, G. H. & Johns, E. W. (1976) *Eur. J. Biochem.* **62**, 461–469
Walker, J. M., Hastings, J. R. B. & Johns, E. W. (1977) *Eur. J. Biochem.* **76**, 461–468
Walker, J. M., Hastings, J. R. B. & Johns, E. W. (1978*a*) *Nature (London)* **271**, 281
Walker, J. M., Parker, B. M. & Johns, E. W. (1978*b*) *Int. J. Peptide Protein Res.* **12**, 269–276
Walker, J. M., Goodwin, G. H. & Johns, E. W. (1979) *FEBS Lett.* **100**, 394–398