*Proc. Natl. Acad. Sci. USA*
Vol. 78, No. 6, pp. 3473–3477, June 1981
Biochemistry

# Amino acid sequence of *Escherichia coli* alkaline phosphatase

(primary structure/serine hydrolase/Edman degradation/secondary structure prediction)

RALPH A. BRADSHAW*, FIORELLA CANCEDDA*†, LOWELL H. ERICSSON‡, PETER A. NEUMANN*§, STEVEN P. PICCOLI*, MILTON J. SCHLESINGER¶, KAREN SHRIEFER*, AND KENNETH A. WALSH‡

*Departments of Biological Chemistry and ¶Microbiology and Immunology, Washington University School of Medicine, St. Louis, Missouri 63110; †Istituto Internazionale di Genetica e Biophysica, Naples, Italy; and ‡Department of Biochemistry, University of Washington, Seattle, Washington 98195

*Communicated by Hans Neurath, March 9, 1981*

**ABSTRACT** **The complete amino acid sequence of the *Escherichia coli* alkaline phosphatase subunit [orthophosphoric-monoester phosphohydrolase (alkaline optimum), EC 3.1.3.1, isozyme 3] has been determined. The monomer contains 449 amino acid residues in a single unglycosylated polypeptide chain having a calculated $M_r$ of 47,029. Isozyme 1 has an additional arginine residue at the $NH_2$ terminus that presumably results from variability in processing of precursor molecules. Sequence data were obtained from both manual and automatic Edman degradation of the tryptic and cyanogen bromide peptides, as well as other peptides derived therefrom. The two disulfide bonds were determined from analyses of the appropriate peptic peptides. This structure confirms earlier reports of the sequence surrounding the active-site serine and both the $NH_2$- and COOH-terminal cyanogen bromide fragments. A secondary structure prediction places nearly half the residues in α-helical segments that have 13% and 16%, respectively, in β-strand and β-turn orientations.**

Alkaline phosphatase [orthophosphoric-monoester phosphohydrolase (alkaline optimum), EC 3.1.3.1] is a metalloenzyme (1) containing two classes of zinc (II) ions (2), as well as magnesium ions (3), that is specified by the *PhoA* gene of *Escherichia coli* (4). The active form of the enzyme is transported to the periplasmic space (5), a process that presumably requires limited proteolysis of a precursor, where it exists as a dimer of identical subunits (6) having a pH optimum of 8.0 (7). Various molecular weights have been reported for the dimer ranging from 80,000 to 100,000 with 85,000–90,000 the most commonly cited (8, 9). As the enzyme is devoid of thiol groups and the subunits are associated by noncovalent bonds, the half-cystinyl residues are in the form of intrachain disulfide bonds (10).

The gross morphological features, including the location and environment of the two $Zn^{2+}$, of the enzyme have been determined by x-ray diffraction to a resolution of 7.7 Å (11). A 3.0 Å electron density map, currently being interpreted, shows a long stretch of α-helix (12). In addition, the number and position of the tyrosine and histidine residues have been examined by NMR spectroscopy (13–15). These analyses have identified histidine residues as the sole ligands for the active-center $Zn^{2+}$ (16).

Alkaline phosphatase occurs in three forms that have been designated isozymes 1, 2, and 3. Although at first dismissed as impurities in the preparation (17), it was subsequently shown that these forms differ by the presence of an $NH_2$-terminal arginine residue on the subunits of isozyme 1 and the absence of the same on isozyme 3, with isozyme 2 representing a heterodimer of the two types of chains (18, 19). The relative proportion of each isozyme depends on the growth conditions of the cells (20). The sequence analysis was performed solely with isozyme 3.

The first sequence information for this enzyme was provided by analyses of the peptides (with 3 and 14 residues) containing the active-site serine (21, 22). This residue is labeled by incubation of the enzyme with orthophosphate at pH 4.0 and 0°C for 10 min (22). Dephosphorylation is dependent on both pH and the metal ion bound in the active site (12). Subsequently, the sequences of the $NH_2$- and COOH-terminal CNBr fragments, containing four and seven residues, respectively, were reported (18). An $NH_2$-terminal sequence analysis of whole enzyme extended the former segment by 15 residues (corresponding to the $NH_2$-terminal arginine plus the first 19 residues of isozyme 3) (17). The identifications of the residues at positions 13 and 15 in that study have been corrected in this work (both amide assignments). In other studies (23), a tryptic peptide derived from a prematurely terminated enzyme produced by an ochre mutant has been placed in the sequence, on the basis of composition, at approximately residues 162–166. Extensive $NH_2$-terminal proteolysis apparently occurred during the isolation of the biosynthetic product. Recently, the structural gene for this enzyme has been cloned into the M13 vector. The sequences of two segments of the DNA have been determined and were identified, by comparison with the protein sequence, as corresponding precisely to residues 140–177 and 363–413 (H. Inouye, W. Barnes, and J. Beckwith, personal communication). Preliminary reports on the progress of the sequence analysis, including a partial structure (9), have appeared previously (24–26).

## SEQUENCE ANALYSIS

As shown in Fig. 1, the principal route for the determination of the primary structure of *E. coli* alkaline phosphatase was via the isolation and characterization of the CNBr fragments. This reaction, always performed on enzyme that had not been reduced and alkylated, produced the nine fragments expected from the eight methionine residues. However, the presence of a CNBr-resistant methionine-threonine bond resulted in the production of a significant amount of a 10th fragment composed of two of the other pieces (CN II and CN Ib). Three intermediate-size fragments (CN III, CN IV, and CN V), containing 82, 49, and 42 residues, were obtained in homogeneous form from the initial gel filtration performed on a Sephadex G-75 column. The sequences of these fragments were mostly determined directly in the automated sequencer with the remaining residues being assigned from manual and automatic Edman degradation of the peptides generated by various enzymatic subdigests. These latter peptides (many of which are not shown in Fig. 1) also served to confirm the direct analyses. Three small frag-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

§ Present address: ANALYTICA, Forchstrasse 138, 8032 Zurich, Switzerland.

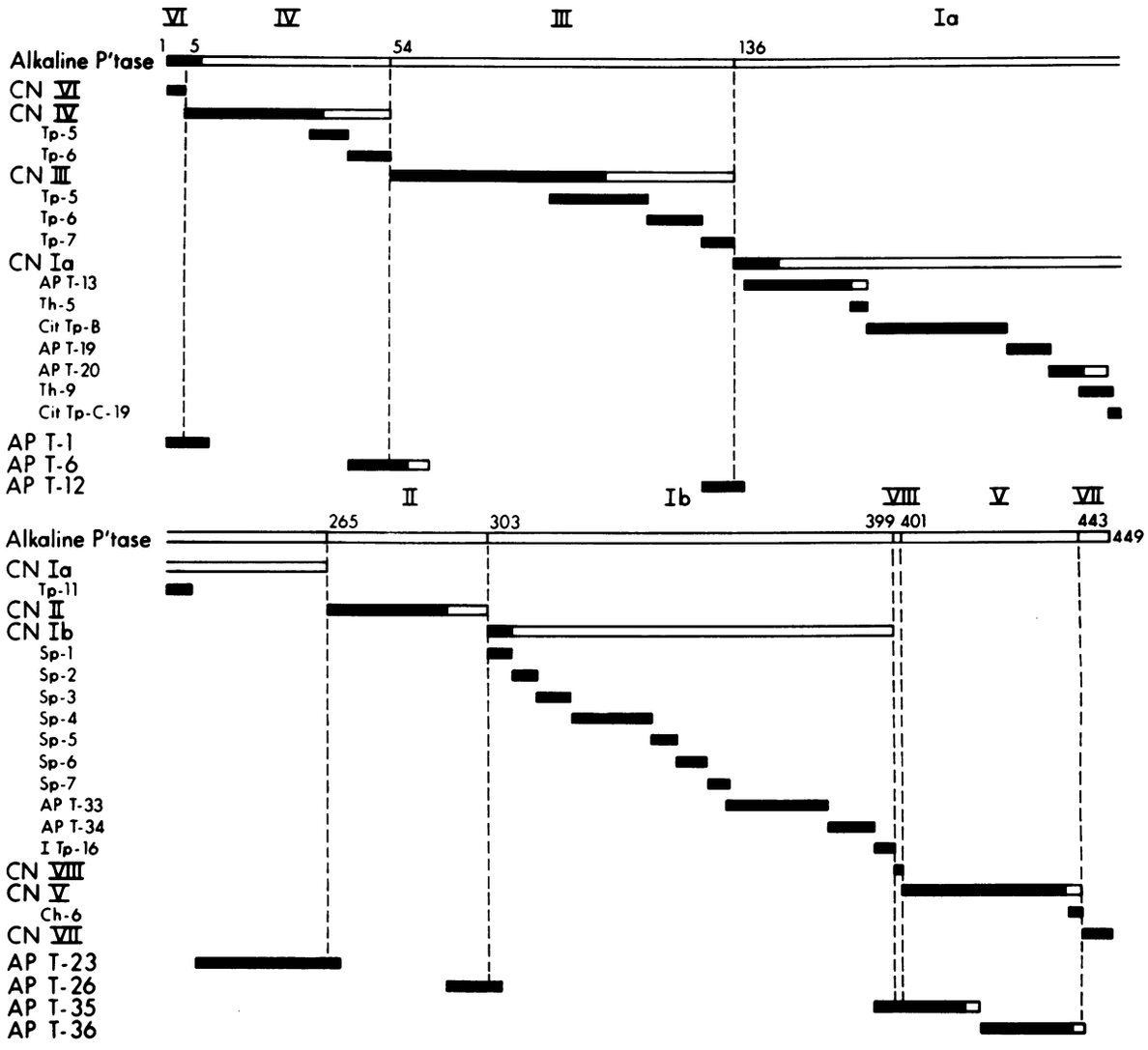


FIG. 1. Schematic representation of the principal peptides used in the sequence analysis of *E. coli* alkaline phosphatase. The upper grouping of each set represents the CNBr fragments and peptides derived therefrom, and the lower set represents the tryptic peptides containing the overlapping methionine sequences derived from whole enzyme. Solid and open bars represent sequenced and unsequenced segments, respectively. Only the minimum peptides used for *residue assignments* are shown. Roman numerals, CNBr fragments; arabic numerals, first residue in each fragment; AP T, tryptic peptide of whole alkaline phosphatase; Cit Tp, tryptic peptide of citraconylated CNBr fragment; Th, thermolytic peptide; Ch, chymotryptic peptide; Sp, *Staphylococcus aureus* V8 protease peptide.

ments (CN VI, CN VII, and CN VIII), which eluted as a single peak near the column volume, were purified on a column of Dowex 50 × 8 (27). Fragments CN VI and CN VII represent the $NH_2$- and COOH-terminal fragments, respectively, that were previously analyzed (18), and CN VIII is a dipeptide, valine-homoserine.

The first fraction to elute from the Sephadex column contains three peptides, CN Ia, CN Ib, and CN II, which comprise about 60% of the molecule. Fragment CN II contains 38 residues and is joined to CN Ib by a disulfide bond. After reduction and alkylation of the initial gel filtration fraction, fragment CN II was separated from the two large fragments, CN Ia and CN Ib, by means of the same Sephadex column. The sequence of fragment CN II was obtained in the fashion described for fragments CN III, CN IV, and CN V. CN Ia and CN Ib were separated by isoelectric focusing and DEAE-cellulose column chromatography. However, these separations have been difficult to reproduce on a preparative scale because of the variable amounts of the partial cleavage product (CN II Ib), and a considerable amount of the sequence data for them was obtained on peptides derived from the mixture of CN Ia and CN Ib (denoted CN I). Of particular value in this regard were the peptides produced by the tryptic hydrolysis of citraconylated CN I. Peptides produced by thermolytic and *S. aureus* V8 protease digests also provided essential information.

Two digests of the whole alkaline phosphatase subunit were required to align the CNBr fragments and determine the disulfide pairing. In the first case, the tryptic peptides of both the *S*-carboxymethyl and *S*-aminoethyl derivatives, containing all of the methionine residues, were isolated by various ion exchange chromatographic methods and provided an unambiguous ordering of the CNBr fragments. Many of the relevant peptides were quite large, necessitating the use of DE-52 cellulose in place of the substituted polystyrenes used in other peptide separations (27). The order of CNBr fragments determined, VI-IV-III-Ia-II-Ib-VIII-V-VII, containing 4, 49, 82, 129, 38, 96, 2, 42, and 7 residues, respectively, confirmed the alignment first predicted by pulse labeling cultures of *E. coli* with [$^{14}$C]-

methionine (24) and subsequently determined independently by a partial CNBr fragmentation technique (28). The second digest of the intact molecule by pepsin yielded two disulfide-containing peptides (not shown in Fig. 1). One, subsequently located in CN Ia, was a single peptide that had the disulfide bond as a short loop (nine intervening residues). The other was formed by two peptides that were severed by performic oxidation and separated by gel filtration. Their location in CN II and CN Ib was determined from their amino acid composition.

## STRUCTURAL AND CATALYTIC FEATURES

The complete covalent structure of the *E. coli* alkaline phosphatase subunit is shown in Fig. 2. The 449 residues produce a polypeptide of $M_r$ 47,029, giving a value for the dimer of 94,058. The calculated average residue $M_r$ of 104.7 is somewhat less than usually found, reflecting the higher concentration of smaller amino acids present. The $M_r$ of the enzyme determined from the covalent structure, although in the range reported from hydrodynamic measurements (8), is larger than the value commonly cited (see ref. 9). This is reflected in the several amino acid compositions that have been reported (6, 29, 30). These data, particularly those of Simpson *et al.* (29), are in excellent agreement with the composition obtained from the sequence when corrected to the true $M_r$. The most serious deviation is the number of tryptophan residues, variously reported as four or five. As only three residues were found in the sequence analysis, the tryptophan content was redetermined spectrophotometrically and found to be 2.84 residues per 11 tyrosine residues (31).

A number of amino acids contribute to the catalytic site of

```
                                                  10                                                  20
Thr-Pro-Glu-Met-Pro-Val-Leu-Glu-Asn-Arg-Ala-Ala-Gln-Gly-Asn-Ile-Thr-Ala-Pro-Gly-

                                                  30                                                  40
Gly-Ala-Arg-Arg-Leu-Thr-Gly-Asp-Gln-Thr-Ala-Ala-Leu-Arg-Asn-Ser-Leu-Ser-Asp-Lys-

                                                  50                                                  60
Pro-Ala-Lys-Asn-Ile-Ile-Leu-Leu-Ile-Gly-Asp-Gly-Met-Gly-Asp-Ser-Glu-Ile-Thr-Ala-

                                                  70                                                  80
Ala-Arg-Asn-Tyr-Ala-Glu-Gly-Ala-Gly-Gly-Phe-Phe-Lys-Gly-Ile-Asp-Ala-Leu-Pro-Leu-

                                                  90                                                 100
Thr-Gly-Gln-Tyr-Thr-His-Tyr-Ala-Leu-Asn-Lys-Lys-Thr-Gly-Lys-Pro-Asp-Tyr-Val-Thr-

                                                 110                                                 120
Asp-Ser-Ala-Ala-Ser-Ala-Thr-Ala-Trp-Ser-Thr-Gly-Val-Lys-Thr-Tyr-Asn-Gly-Ala-Leu-

                                                 130                                                 140
Gly-Val-Asp-Ile-His-Glu-Lys-Asp-His-Pro-Thr-Ile-Leu-Glu-Met-Ala-Lys-Ala-Ala-Gly-

                                                 150                                                 160
Leu-Ala-Thr-Gly-Asn-Val-Ser-Thr-Ala-Glu-Leu-Gln-Asp-Ala-Thr-Pro-Ala-Ala-Leu-Val-

                                                 170                                                 180
Ala-His-Val-Thr-Ser-Arg-Lys-Cys-Tyr-Gly-Pro-Ser-Ala-Thr-Ser-Gln-Lys-Cys-Pro-Gly-

                                                 190                                                 200
Asn-Ala-Leu-Glu-Lys-Gly-Gly-Lys-Gly-Ser-Ile-Thr-Glu-Gln-Leu-Leu-Asn-Ala-Arg-Ala-

                                                 210                                                 220
Asp-Val-Thr-Leu-Gly-Gly-Gly-Ala-Lys-Thr-Phe-Ala-Glu-Thr-Ala-Thr-Ala-Gly-Glu-Trp-

                                                 230                                                 240
Gln-Gly-Lys-Thr-Leu-Arg-Glu-Gln-Ala-Gln-Ala-Arg-Gly-Tyr-Gln-Leu-Val-Ser-Asp-Ala-

                                                 250                                                 260
Ala-Ser-Leu-Asn-Ser-Val-Thr-Glu-Ala-Asn-Gln-Gln-Lys-Pro-Leu-Leu-Gly-Leu-Phe-Ala-

                                                 270                                                 280
Asp-Gly-Asn-Met-Pro-Val-Arg-Trp-Leu-Gly-Pro-Lys-Ala-Thr-Tyr-His-Gly-Asn-Ile-Asp-

                                                 290                                                 300
Lys-Pro-Ala-Val-Thr-Cys-Thr-Pro-Asn-Pro-Gln-Arg-Asn-Asp-Ser-Val-Pro-Thr-Leu-Ala-

                                                 310                                                 320
Gln-Met-Thr-Asp-Lys-Ala-Ile-Glu-Leu-Leu-Ser-Lys-Asn-Glu-Lys-Gly-Phe-Phe-Leu-Gln-

                                                 330                                                 340
Val-Glu-Gly-Ala-Ser-Ile-Asp-Lys-Gln-Asp-His-Ala-Ala-Asn-Pro-Cys-Gly-Gln-Ile-Gly-

                                                 350                                                 360
Glu-Thr-Val-Asp-Leu-Asp-Glu-Ala-Val-Gln-Arg-Ala-Leu-Glu-Phe-Ala-Lys-Lys-Glu-Gly-

                                                 370                                                 380
Asn-Thr-Leu-Val-Ile-Val-Thr-Ala-Asp-His-Ala-His-Ala-Ser-Gln-Ile-Val-Ala-Pro-Asp-

                                                 390                                                 400
Thr-Lys-Ala-Pro-Gly-Leu-Thr-Gln-Ala-Leu-Asn-Thr-Lys-Asp-Gly-Ala-Val-Met-Val-Met-

                                                 410                                                 420
Ser-Tyr-Gly-Asn-Ser-Glu-Glu-Asp-Ser-Gln-Glu-His-Thr-Gly-Ser-Gln-Leu-Arg-Ile-Ala-

                                                 430                                                 440
Ala-Tyr-Gly-Pro-His-Ala-Ala-Asn-Val-Val-Gly-Leu-Thr-Asp-Gln-Thr-Asp-Leu-Phe-Tyr-

                                                 449
Thr-Met-Lys-Ala-Ala-Leu-Gly-Leu-Lys
```

FIG. 2. The amino acid sequence of *E. coli* alkaline phosphatase (isozyme 3).

Table 1. Active-site sequences of serine hydrolases

| | Sequence | Ref. |
|---|---|---|
| Alkaline phosphatase | | |
| *E. coli* | Thr-Gly-Lys-Pro-Asp-Tyr-Val-Thr-Asp-**Ser**-Ala-Ala-Ser-Ala | (20) |
| *Serratia marcescens* | Thr/Ser-Asp-**Ser**-Ala | (33) |
| Calf intestine | Asp-**Ser**-Ala | (34) |
| Butyryl cholinesterase | Phe-Gly-Gly-Glu-**Ser**-Ala-Gly | (35) |
| Chymotrypsin | Ser-Gly-Val-Ser-Ser-Cys-Met-Gly-Asp-**Ser**-Gly-Gly-Pro-Leu | (36) |
| Liver aliesterase | Gly-Glu-**Ser**-Ala-Gly-Gly | (37) |
| Subtilisin | Thr-**Ser**-Met-Ala | (38) |
| Trypsin | Gly-Gly-Lys-Asp-Ser-Cys-Gln-Gly-Asp-**Ser**-Gly-Gly-Pro-Val | (39) |

Boldface type indicates active-site serine.

the enzyme. In addition to the serine residue, identified by phosphorylation (21, 22), there appears to be an arginine residue (32) and an atom of zinc (II) coordinated by four histidine ligands, as judged by NMR experiments on $^{113}Cd^{2+}$-substituted derivatives of the enzyme (16). The second gram atom of zinc is bound to each subunit outside of the active site, apparently serving in a structural role, and has only one bound histidine (16), the other ligands being still unidentified. The $Mg^{2+}$ is not coordinated by histidine residues. Neither the arginine nor any of the metal-binding residues have been identified in the sequence. However, the His-Ala-His sequence at positions 370–372 is an attractive possibility for two of the zinc ligands at the catalytic center.

The reactive serine residue occurs at position 102. As shown in Table 1, the amino acids immediately adjacent to this residue are similar to those found in many other serine hydrolases. In fact, a sequence of Asp/Glu-Ser-Ala/Gly is clearly characteristic of such enzymes occurring in all those listed but subtilisin, regardless of origin or function.

A predictive analysis of the secondary structural elements of *E. coli* alkaline phosphatase by the formulations of Chou and Fasman (40) suggests the three-dimensional structure will be characterized by a considerable amount of $\alpha$-helix. Approximately 22 segments, comprising about 200 residues ($\approx$45%), were identified by this analysis. Interestingly, optical rotatory dispersion measurements suggest 40% $\alpha$-helix (41). The longest helix predicted was 21 residues in the segment extending just beyond the last half-cystinyl residue (which itself is suggested to reside in a $\beta$-turn). Only 13% of the residues were found in $\beta$-sheets, all in relatively short stretches. There were 18 $\beta$-turns, using 72 residues (16%), fairly evenly distributed throughout the molecule. The first disulfide bond is contained in a random (or irregular) segment, and the second connects a short segment of $\beta$-sheet with a $\beta$-turn. The overall correctness of these predictions, as well as the other structural features of the enzyme, will require completion of the interpretation of the electron density maps by Wyckoff and his colleagues. The structure of the enzyme so produced should provide the basis for defining the functional characteristics in molecular terms and present an excellent opportunity to develop further the relationship of structure and function through characterization of the many mutant forms of the enzyme already identified (42).

We wish to thank Dr. George Long for performing the Chou–Fasman analysis and Drs. Wayne Barnes and H. Inouye for providing unpublished data. Prof. Hans Neurath kindly provided facilities for some of the analyses. This work was supported by research grants from the National Science Foundation (GB-27613) and the National Institutes of Health (AM 13362 and GM 15731).

1. Plocke, D. J., Levinthal, C. & Vallee, B. L. (1962) *Biochemistry* **1**, 373–378.
2. Simpson, R. T. & Vallee, B. L. (1968) *Biochemistry* **7**, 4343–4350.
3. Bosron, W. F., Anderson, R. A., Falk, M. C., Kennedy, F. S. & Vallee, B. L. (1977) *Biochemistry* **16**, 610–614.
4. Garen, A., Levinthal, C. & Rothman, F. (1961) *J. Chem. Phys.* **58**, 1068–1071.
5. Melamy, M. H. & Horecker, B. L. (1964) *Biochemistry* **3**, 1889–1897.
6. Rothman, F. & Byrne, R. (1963) *J. Mol. Biol.* **6**, 330–340.
7. Garen, A. & Levinthal, C. (1960) *Biochim. Biophys. Acta* **38**, 470–483.
8. Applebury, M. L. & Coleman, J. E. (1969) *J. Biol. Chem.* **244**, 308–318.
9. McComb, R. B., Bowers, G. N., Jr. & Posen, S. (1979) *Alkaline Phosphatase* (Plenum, New York), pp. 189–228.
10. Schlesinger, M. J. & Barrett, K. (1965) *J. Biol. Chem.* **240**, 4284–4292.
11. Knox, J. R. & Wyckoff, H. W. (1973) *J. Mol. Biol.* **74**, 533–545.
12. Coleman, J. E., Chlebowski, J. F., Otvos, J. D., Uiterkamp, A. J. M. S. & Armitage, I. (1978) *Trans. Am. Crystallogr. Assoc.* **14**, 17–55.
13. Sykes, B. D., Weingarten, H. I. & Schlesinger, M. J. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 469–473.
14. Hull, W. E. & Sykes, B. D. (1976) *Biochemistry* **15**, 1535–1544.
15. Otvos, J. D. & Browne, D. T. (1980) *Biochemistry* **19**, 4011–4021.
16. Otvos, J. D. & Armitage, I. M. (1980) *Biochemistry* **19**, 4021–4030.
17. Bridgen, J. & Secher, D. S. (1973) *FEBS Lett.* **29**, 55–57.
18. Kelley, P. M., Neumann, P. A., Shriefer, K., Cancedda, F., Schlesinger, M. J. & Bradshaw, R. A. (1973) *Biochemistry* **12**, 3499–3503.
19. Schlesinger, M. J., Block, W. & Kelley, P. M. (1975) in *Isozymes I: Molecular Structure*, ed. Markert, C. (Academic, New York), pp. 333–342.
20. Schlesinger, M. J. & Anderson, L. (1968) *Ann. N.Y. Acad. Sci.* **151**, 159–170.
21. Milstein, C. (1963) *Biochim. Biophys. Acta* **67**, 171–172.
22. Schwartz, J. H., Crestfield, A. & Lipmann, F. (1963) *Proc. Natl. Acad. Sci. USA* **49**, 722–729.
23. Natori, S. & Garen, A. (1970) *J. Mol. Biol.* **79**, 577–588.
24. Neumann, P. A., Bradshaw, R. A. & Schlesinger, M. J. (1971) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **30**, 1242.
25. Bradshaw, R. A., Neumann, P. A., Cancedda, F., Schriefer, K., Hecht, J. D. & Schlesinger, M. J. (1975) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **34**, 630.
26. Piccoli, S. P., Cancedda, F., Neumann, P. A., Shriefer, K., Schlesinger, M. J. & Bradshaw, R. A. (1980) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **39**, 1997.
27. Bradshaw, R. A., Bates, O. J. & Benson, J. R. (1980) *J. Chromatogr.* **187**, 27–33.
28. Nikodem, V. & Fresco, J. R. (1979) *Anal. Biochem.* **97**, 382–386.
29. Simpson, R. T., Vallee, B. L. & Tait, G. H. (1968) *Biochemistry* **7**, 4336–4342.
30. Lazdunski, C. & Lazdunski, L. (1967) *Biochim. Biophys. Acta* **147**, 280–288.
31. Edelhoch, H. (1967) *Biochemistry* **6**, 1948–1954.
32. Daemen, F. J. M. & Riordan, J. F. (1974) *Biochemistry* **13**, 2865–2877.

33. Zwaig, M. & Milstein, C. (1964) *Biochem. J.* **92,** 421–422.
34. Engstrom, L. (1964) *Biochim. Biophys. Acta* **92,** 79–84.
35. Jansy, H. S., Brons, D. & Warringa, M. G. P. J. (1959) *Biochim. Biophys. Acta* **34,** 573–575.
36. Hartley, B. S. (1962) *Brookhaven Symp. Biol.* **15,** 85–100.
37. Jansy, H. S., Posthumus, C. H. & Cohen, J. A. (1959) *Biochim. Biophys. Acta* **33,** 396–403.
38. Sanger, F. & Shaw, D. C. (1960) *Nature (London)* **187,** 872–873.
39. Walsh, K. A. (1970) *Methods Enzymol.* **19,** 41–63.
40. Chou, P. Y. & Fasman, G. D. (1978) in *Advances in Enzymology,* ed. Meister, A. (Wiley, New York), Vol. 47, pp. 45–148.
41. Reynolds, J. A. & Schlesinger, M. J. (1967) *Biochemistry* **6,** 3552–3559.
42. Schlesinger, M. J. (1974) *Biochem. Soc. Trans.* **2,** 827–831.