Proc. Natl. Acad. Sci. USA
Vol. 76, No. 7, pp. 3218–3222, July 1979
Biochemistry

Favin versus concanavalin A: Circularly permuted amino acid sequences

(lectins/three-dimensional structure/sequence homology/carbohydrate)

BRUCE A. CUNNINGHAM, JOHN J. HEMPERLY, THOMAS P. HOPP, AND GERALD M. EDELMAN

The Rockefeller University, 1230 York Avenue, New York, New York 10021

Contributed by Gerald M. Edelman, April 16, 1979

ABSTRACT We have determined the tentative amino acid sequence of the β chain (M_r 20,000) of the lectin favin. In previous studies, we have shown that the α chain (M_r 5600) of this lectin is homologous to a region in the middle of the concanavalin A (Con A) sequence (residues 70–119). Now we present evidence that the β chain is homologous to two discrete segments of Con A. The homology begins at residue 120 of Con A, extends to the COOH terminus (residue 237) and continues without interruption through the NH_2-terminal 69 residues of Con A. Together, the α and β chains of favin account for a polypeptide chain equivalent in size to that of Con A. The comparison of the two proteins thus reveals a circular permutation of extensive homologous sequences. The favin molecule contains residues identical to many of the residues postulated to be involved in sugar binding by Con A, and contains all of the direct metal ligands as well as residues homologous to most of the residues that form the β-pleated sheets of Con A. These homologies suggest that the three-dimensional structures of the two lectins are likely to be very similar. Moreover, favin appears to be even more closely related in primary structure and sugar specificity to the lectins from pea and lentil, raising the possibility that all of these lectins may have structures that resemble Con A. Some of these similarities may also extend to the lectins from soybean, peanut, and red kidney bean, which have different sugar specificities but share sequence homologies with the favin β chain.

Plant lectins bind various carbohydrates and interact with cells via the carbohydrate moieties of glycoproteins and glycolipids on the cell surface. Lectin binding induces diverse effects in animal cells, including agglutination, mitogenesis, and alteration of receptor mobility (1). Various lectins with different sugar specificities have been isolated from plant sources (2), but the complete amino acid sequence and the three-dimensional structure have been determined (3, 4) only for concanavalin A (Con A). This lectin is arranged as a tetramer composed of four identical polypeptide chains, each containing 237 residues (5–8).

Various lectins from other seeds show some sequence homology with Con A but differ from Con A in their subunit structure. This group includes the lentil (9) and pea lectins (10, 11), which have the same sugar specificities as Con A but, unlike Con A, are each composed of two types of chains (α, M_r 5800–5900 and β, M_r approximately 17,000). Other lectins in this group include those from peanut and soybean (12) and two lectins (R-PHA and L-PHA) from the red kidney bean (13), which differ in sugar specificity and contain only one (presumably β-like) chain. The α chains of the lentil and pea lectins are homologous to each other and to residues 72–122 of Con A (9–11). The NH_2-terminal regions of the β chains of these lectins and the β-like chains of the other lectins are all homologous to each other and to a region beginning at residue 123 of Con A (12).

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.

We have recently (14) characterized the lectin favin (15) from *Vicia faba*. This protein, which binds glucose and mannose, is composed of α and β chains that closely resemble those of the lentil and pea lectins, and it shares many biological activities with these lectins and Con A. At physiological pH, favin has a M_r of 51,000 and is therefore probably composed of two α and two β chains. Favin also differs from Con A in that it contains covalently bound carbohydrate. We have previously reported (14) the amino acid sequence of the α chain of favin; it is closely homologous to residues 72–120 of Con A.

All of these findings suggested that the six lectins considered above lack regions corresponding to residues 1–71 of Con A. This conclusion is somewhat unsatisfying because this portion of the Con A molecule is a compact structural unit that contains most of the ligands for the metal atoms (7) and some residues presumed to be involved in carbohydrate binding (16, 17).

The present determination of the tentative amino acid sequence of the β chain of favin has resolved the issue at least for this protein. Comparison with Con A indicates that the COOH-terminal portion of the β chain is homologous to residues 1–69 of Con A, suggesting that the carbohydrate-binding and metal-binding sites of Con A are also present in favin and that the two proteins are evolutionarily related. This appears to be the first example of a circular permutation in the arrangement of homologous amino acid sequences of proteins having similar specificity and presumably similar functions.

MATERIALS AND METHODS

Favin, the α chain, and the β (plus β') chain were prepared as described (14). Digestions with trypsin (Worthington), *Staphylococcus aureus* protease (Miles), thermolysin (Calbiochem), and chymotrypsin (Worthington) were performed in 1% NH_4HCO_3; digestions with pepsin (Worthington) were carried out in 5% formic acid. The β chain was also treated with citraconic anhydride in dilute sodium hydroxide (18) prior to digestion with trypsin. Cleavage with hydroxylamine was carried out as described by Bornstein and Balian (19).

Peptides were purified by gel filtration on Bio-Gel P-6 and Sephadex G100 and by electrophoresis on paper at pH 4.7 (20) plus chromatography on paper in 1-butanol/pyridine/acetic acid/water (15:10:3:12) (21).

Amino acid analysis was carried out on a Beckman 121M amino acid analyzer. Peptide sequences were determined on a Beckman 890C Sequenator, equipped with a Sequemat P-6 automatic converter using a Quadrol program and Polybrene as a carrier, as described (14). Glycopeptides were detected by the phenol/sulfuric acid method (22) and quantitative sugar analyses were performed by gas chromatography of the aldol acetate derivatives (23, 24).

Abbreviation: Con A, concanavalin A.

RESULTS

The amino acid sequence of the β chain of favin was determined by the isolation and characterization of peptides T1–T12 (Fig. 1) obtained after digestion with trypsin. The order of these peptides was initially established by direct sequence determination of the intact chain and by isolation and characterization of peptides obtained from digests of the chain with chymotrypsin. This order was confirmed and the remaining sequence was determined by characterizing peptides obtained after cleavage of the intact β chain and the tryptic peptides both chemically and with other enzymes.

Tryptic peptides from the β chain were initially fractionated on a column of Bio-Gel P-6 in 0.015 M NH_4OH/10% 1-propanol. Larger peptides appearing in the void volume were purified further by gel filtration on a column (1.8 × 100 cm) of Sephadex G100 in 6 M guanidine to give peptides T7, T9, and T12. The remaining peptides were isolated from the material in the included volume of the P-6 column by high-voltage electrophoresis and chromatography on paper.

Peptide T7 was too large for complete direct sequence analysis and peptides T9 and T12 each contained Asn-Gly sequences that stopped adequate sequence analysis after 21 cycles in T9 and 6 cycles in T12. The sequence determinations of T9, T12, and all but two short segments (residues 85–87 and 102–107) in T7 were completed after cleavage of these peptides and the intact chain chemically and with other enzymes as outlined in Fig. 1.

Of the tryptic peptides, only T12 gave a strong positive reaction for carbohydrate with phenol/sulfuric acid; peptide T9 gave a weakly positive reaction with this reagent, but we assume that this was due to small amounts of contaminating T12. Gas chromatographic analysis indicated that T12 contained all of the mannose and glucosamine in the intact chain. Sequence analyses of peptide T12-SA1 gave no detectable phenylthiohydantoin-amino acid at position 171, suggesting that this position contained the covalently bound carbohydrate. The amino acid composition of peptide T12-C2 indicated that this residue is aspartic acid or asparagine and sequence data indicated that the next amino acid is alanine. These results are consistent with earlier observations (25) that the glycopeptide of favin contains aspartic acid or asparagine and alanine. Peptide T9 also contains an Asn-Ala sequence, but the asparaginyl residue was obtained in high yield in the sequence analysis of T9, suggesting that there is no significant amount of carbohydrate attached to this residue.

The amino acid sequences of the α and β chains of favin are

FIG. 1. Amino acid sequence analysis of the favin β chain. The solid bars denote peptides isolated after enzymatic digestion or chemical degradation: T, trypsin; C, chymotrypsin; P, pepsin; SA, *Staphylococcus aureus* protease; CT, trypsin digestion after treatment of β chain with citraconic anhydride; and HA, peptide cleavage with hydroxylamine. Arrows denote automated sequence analysis: ⇀, directly on the intact chain; →, on purified peptides; —, residue not determined at this position; ⩘, NH_2-terminal residue by the dansyl technique.

compared with the sequence of Con A in Fig. 2. By inspection, and in accord with earlier comparisons of the α chains of lentil and pea lectins with Con A (9–12), the α chain of favin is placed so that it is aligned with residues 70–119 of Con A. Similarly, the β chain of favin is arranged so that it is aligned beginning with residue 120 and extending to the COOH terminus of Con A (residue 237). The remaining sequence of the favin β chain is aligned with residues 1–69 of Con A. This alignment is summarized schematically in Fig. 3.

When aligned in this fashion, identical residues appear in nearly 40% of the positions (Fig. 2). To refine the initial alignments [which had placed the α chain at Con A position 72 (14)] and to obtain more quantitative comparisons, the metric analysis program (26, 27) was used to determine the number of base-pair changes needed to convert one sequence into the other and to determine the statistical significance of the homology observed. These comparisons showed that only 160 base-pair changes with 19 gaps (assigned at 2.01 base changes per gap) were required to interconvert the two sequences. They also indicated that the homology was highly significant in that it gave the fewest base-pair changes of at least 1000 randomly permuted sequences having the same composition. Alternative alignments of the α plus β chain with Con A, the α chain with Con A, and the β chain with Con A all gave either much less or no significant homology.

DISCUSSION

Favin and Con A have similar sugar specificities and similar biological activities (15, 25). Con A, however, is composed of a single polypeptide chain (3), whereas favin contains two chains, α and β (14, 15). If the α and β chains of favin are arranged sequentially and aligned with Con A in a circle (Fig. 3),

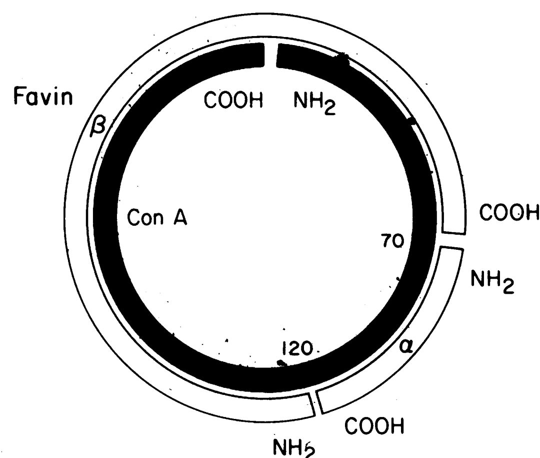

FIG. 3. Schematic drawing of the alignment of favin α and β chains (open bar) with Con A (solid bar) showing the circular permutation that gave maximum homology between the two sequences as shown in Fig. 2.

the two proteins are strikingly similar throughout their sequences with the COOH-terminal portion of the β chain of favin resembling the NH_2-terminal portion of Con A (Fig. 2). These results suggest that the two proteins are evolutionarily related but that an unusual genetic event has resulted in permutation of homologous sequences.

The homology between favin and Con A is so close that it is likely that the three-dimensional structures of the two proteins will be found to be similar. For example, of the 57 residues in the central β-pleated sheet of Con A (8), 36 are identical in favin, and 22 of the 64 residues in the back β-sheet of Con A (8)

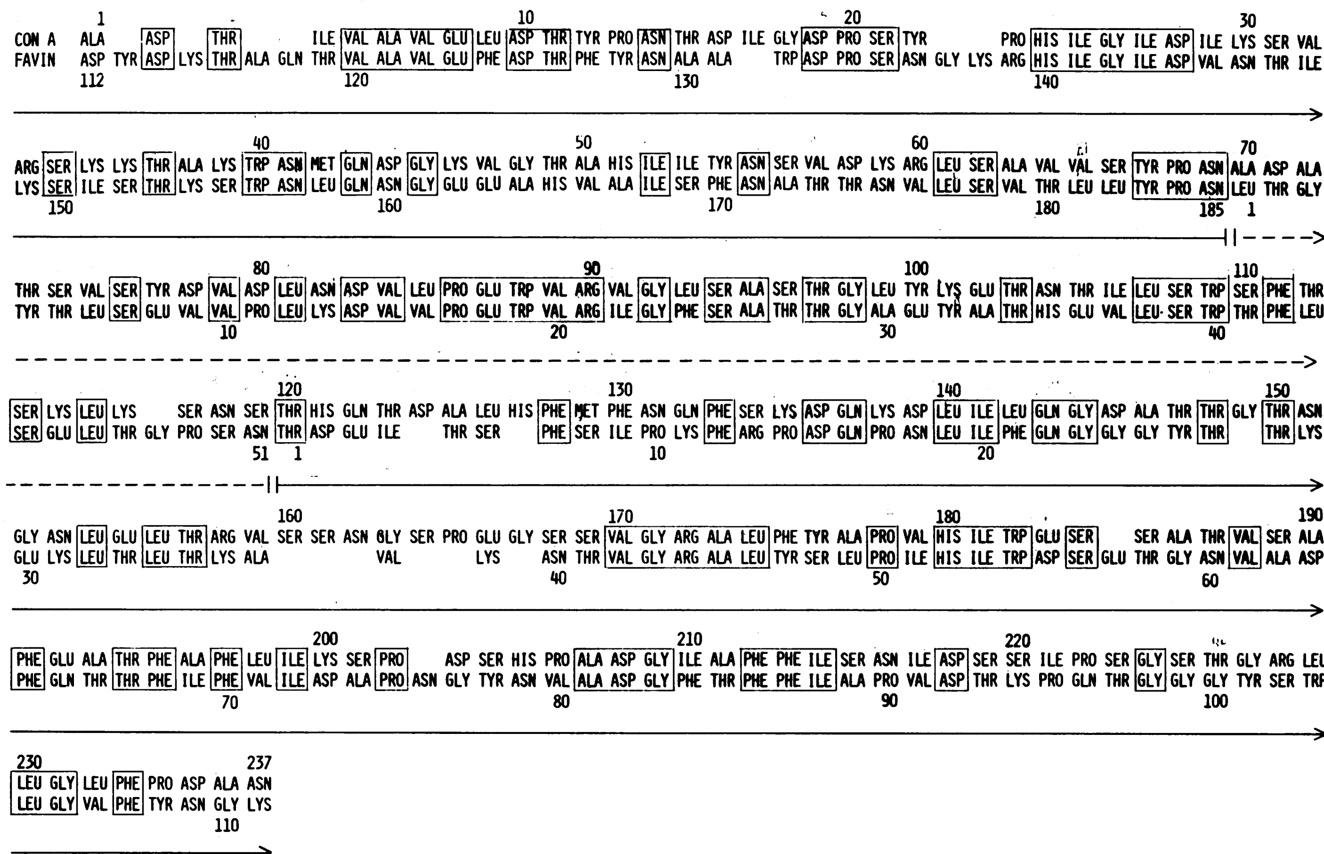

FIG. 2. Comparison of the amino acid sequences of Con A (top line in each row) and favin (bottom line in each row). Gaps were inserted in both sequences to maximize homology. Boxes enclose identical residues. - - -, Favin α chain; —, favin β chain. The α chain sequence is from ref. 14.

are conserved in favin. Of the 18 residues that might be involved in carbohydrate binding in Con A (16), 6 are identical in favin. Moreover, all of the direct metal ligands (7) in Con A are conserved in the favin sequence; preliminary analyses of the metals in favin indicated that it contains stoichiometric amounts of Mn^{2+} and Ca^{2+} as does Con A.

The most striking feature in the homology between favin and Con A is the circular permutation of the two sequences (Fig. 3). This poses the problem of how to generate an amino acid sequence BCA from sequence ABC where each of the letters refers to a fairly large stretch of sequence. A number of possible mechanisms that could generate circularly permuted sequences in evolutionarily related proteins can be imagined. We consider briefly here only those that operate at the level of the gene itself. Clearly, posttranscriptional processes involving splicing could also lead to permuted sequences. Although posttranscriptional mechanisms are feasible, they have no precedent at present; in any case, they are subject to experimental verification.

The most conventional mechanism at the DNA level would involve fusion of tandemly duplicated genes occurring during evolution. This possibility would be supported if it were found that lectins are members of isotypical protein families, indicating that there are preexisting gene duplications in the genome. The absence of such isotypes would not, however, rule out such a fusion. As shown in Fig. 4 *upper*, several independent deletions or mutations leading to failure of expression could lead to sequence permutation. It is also possible that some of these mutational events could be mimicked by posttranslational cleavage of precursor polypeptide chains. A simpler mechanism (Fig. 4 *lower*) would involve a single excision and reinsertion of a nucleotide sequence coding for a considerable part or all of the polypeptide chain or even of a promotor region after gene duplication. So far, this mechanism lacks a precedent within a structural gene of higher plants but transposable regulatory elements of the kind suggested by the work of McClintock (28) provide some basis for such an event. These examples are all

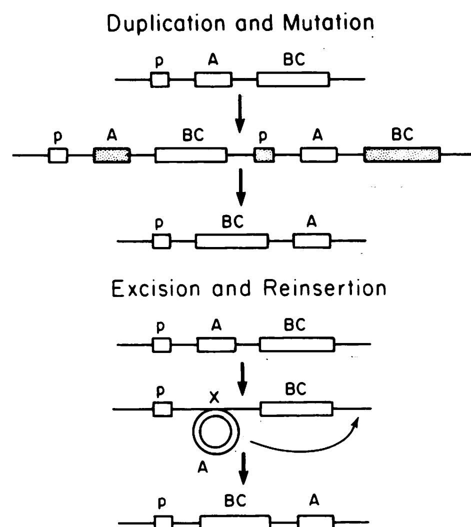

FIG. 4. Diagrams illustrating two possible modes of sequence permutation occurring in DNA. (*Upper*) Duplication and mutation. After duplication, deletions or mutations leading to failure of expression (shown, for example, by shaded areas) could result in a sequence that is permuted. Not all of these are necessary; some could be mimicked by posttranslational cleavage of a precursor polypeptide chain. (*Lower*) Excision and reinsertion. Only one possible excision is illustrated; a larger stretch of sequence could also be excised and reinserted with similar results. □, Coding sequences; —, introns (placed only for completeness); p, promotor or signal sequences; X, crossing-over.

given for conversion of some evolutionary precursor of Con A to that of favin. Obviously, with minor modifications, the same arguments would apply to a precursor of favin if it is assumed that this arose first in evolution.

None of these mechanisms would be drastically altered if the α and β chains of favin were specified by separate genes or if they were specified by a single gene with posttranscriptional modification. The fact that some molecules of Con A are found to be cleaved (29) at the position corresponding to the beginning of the β chain of favin suggests that some posttranslational events may in fact be involved in the biosynthesis of favin. What is needed to discriminate definitively among these various possibilities is gene or messenger isolation, cloning, and polynucleotide sequence determination.

Favin differs from Con A in that it contains covalently bound carbohydrate (15, 25). Our initial studies suggest that the carbohydrate, which contains mannose and glucosamine (25), is localized on the COOH-terminal tryptic peptide (T12) of the β chain. The composition and location of the carbohydrate on an asparaginyl residue followed by an alanine is in accord with the results of Allen *et al.* (25), who isolated a glycopeptide containing asparagine and alanine from a Pronase digest of whole favin.

A major difference between Con A and favin is that Con A is a tetramer (103,000 daltons) (30) of a single peptide chain, whereas favin is a dimer (51,000 daltons) (15) of α and β chain pairs—i.e., $(\alpha\beta)_2$. On the assumption that the three-dimensional structure of favin will be found to be similar to that of Con A, we have compared residues in the contact regions (8). The comparisons suggest that monomer–monomer contacts for dimer formation are probably conserved in favin but that dimer–dimer contacts are not. It is notable that residues 1–69 of Con A which correspond to the permuted sequence form a compactly folded unit comprising the region of the two β structures near the metal binding loop and including that loop. This unit might represent the relic of an ancient metal-binding protein, the gene for which may have been incorporated into an ancestral gene for Con A or favin.

The comparison of other lectins to favin and Con A suggests that a number of these proteins will be closely related to Con A in primary and three-dimensional structure. Comparison of the α chains of favin, lentil, and pea lectins (9–11, 14) indicate that these molecules are more closely related to each other than any one of them is to Con A. This observation raises the possibility that all of these lectins may be related to Con A by a circular permutation of homologous sequences; it is not clear, however, whether they represent a group of molecules that evolved prior to or subsequent to the divergence of *Canavalia* from other legumes. All three of these lectins probably have features of their three-dimensional structures in common with Con A and it may therefore be possible to solve at least some of the three-dimensional structures by molecular replacement methods. Completion of the amino acid sequences of the homolgous lectins from peanut, soybean, and red kidney bean may reveal that, like favin, they have a relationship (circular permutation) to Con A and also have a region comparable to the favin α chain at their COOH termini. It will be of particular interest to determine whether the sugar-binding sites of these lectins, which have different specificities than Con A, are totally different from that of Con A, or whether they are homologous to the sugar-binding site of Con A with only slight variations in specific residues. In addition, the structures of these and other lectins may provide a useful biochemical basis for classifying the leguminosae and describing their evolutionary relationships.

We are grateful to Dr. Reinhard Mailhammer for carrying out the sugar analyses, to Dr. Bruce Erickson for help with the computer programs for comparing sequences, and to Drs. George Reeke and Joseph Becker for valuable discussion. We thank Ms. Kerri Ann Jones, Ms. Nancy Dunn, and Ms. Dorothy Gamarra for excellent technical assistance. This work was supported by Grants AI-11378, AI-09273, and CA-19040 from the U.S. Public Health Service and PCM77-09377 from the National Science Foundation. B.A.C. is the recipient of a Career Scientist Award from the Irma T. Hirschl Trust. T.P.H. is a recipient of a postdoctoral fellowship from the Cancer Research Institute.

1. Edelman, G. M. (1976) *Science* **192,** 218–226.
2. Lis, H. & Sharon, N. (1977) in *The Antigens*, ed. Sela, M. (Academic, New York), Vol. 4, pp. 429–529.
3. Edelman, G. M., Cunningham, B. A., Reeke, G. N., Jr., Becker, J. W., Waxdal, M. J. & Wang, J. L. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 2580–2584.
4. Hardman, K. D. & Ainsworth, C. F. (1972) *Biochemistry* **11,** 4910–4919.
5. Wang, J. L., Cunningham, B. A., Waxdal, M. J. & Edelman, G. M. (1975) *J. Biol. Chem.* **250,** 1490–1502.
6. Cunningham, B. A., Wang, J. L., Waxdal, M. J. & Edelman, G. M. (1975) *J. Biol. Chem.* **250,** 1503–1512.
7. Becker, J. W., Reeke, G. N., Jr., Wang, J. L., Cunningham, B. A. & Edelman, G. M. (1975) *J. Biol. Chem.* **250,** 1513–1524.
8. Reeke, G. N., Jr., Becker, J. W. & Edelman, G. M. (1975) *J. Biol. Chem.* **250,** 1525–1547.
9. Foriers, A., de Neve, R., Kanarek, L. & Strosberg, A. D. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 1136–1139.
10. Van Driessche, E., Foriers, A., Strosberg, A. D. & Kanarek, L. (1976) *FEBS Lett.* **71,** 220–222.
11. Richardson, C., Behnke, W. D., Freisheim, J. H. & Blumenthal, K. M. (1978) *Biochim. Biophys. Acta* **537,** 310–319.
12. Foriers, A., Wuilmart, C., Sharon, N. & Strosberg, A. D. (1977) *Biochim. Biophys. Res. Commun.* **75,** 980–986.
13. Miller, B. J., Hsu, R., Heinrikson, R. & Yachnin, S. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 1388–1391.
14. Hemperly, J. J., Hopp, T. P., Becker, J. W. & Cunningham, B. A. (1979) *J. Biol. Chem.*, in press.
15. Wang, J. L., Becker, J. W., Reeke, G. N., Jr. & Edelman, G. M. (1974) *J. Mol. Biol.* **88,** 259–262.
16. Becker, J. W., Reeke, G. N., Jr., Cunningham, B. A. & Edelman, G. M. (1976) *Nature (London)* **259,** 406–409.
17. Hardman, K. D. & Ainsworth, C. F. (1976) *Biochemistry* **15,** 1120–1128.
18. Atassi, M. Z. & Habeeb, A. F. S. A. (1972) *Methods Enzymol.* **25,** 546–553.
19. Bornstein, P. & Balian, G. (1977) *Methods Enzymol.* **47,** 132–145.
20. Schwarz, J. H. & Edelman, G. M. (1963) *J. Exp. Med.* **118,** 41–54.
21. Milstein, C., Milstein, C. P. & Feinstein, A. (1969) *Nature (London)* **221,** 151–154.
22. DuBois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A. & Smith, F. (1956) *Anal. Chem.* **28,** 350–356.
23. Porter, W. H. (1975) *Anal. Biochem.* **63,** 27–43.
24. Glick, M. C. (1974) in *Methods in Membrane Biology*, ed. Korn, E. D. (Plenum, New York), Vol. 2, pp. 157–204.
25. Allen, A. K., Desai, N. N. & Neuberger, A. (1976) *Biochem. J.* **155,** 127–135.
26. Sellers, P. H. (1974) *J. Appl. Math.* **26,** 787–793.
27. Erickson, B. W., Fulk, M. A. & Sellers, P. H. (1979) in *XIth International Congress of Biochemistry*, Toronto, Canada, in press (abstr.).
28. McClintock, B. (1965) *Brookhaven Symp. Biol.* **18,** 162–184.
29. Wang, J. L., Cunningham, B. A. & Edelman, G. M. (1971) *Proc. Natl. Acad. Sci. USA* **68,** 1130–1134.
30. Kalb, A. J. & Lustig, A. (1968) *Biochim. Biophys. Acta* **168,** 366–367.