*Proc. Natl. Acad. Sci. USA*
Vol. 79, pp. 6852–6855, November 1982
Biochemistry

# Exon cloning: Immunoenzymatic identification of exons of the chicken lysozyme gene

(antigenic determinants/*lac* operon/molecular cloning)

ULRICH RÜTHER, MICHAEL KOENEN, ALBRECHT E. SIPPEL, AND BENNO MÜLLER-HILL

Institut für Genetik der Universität zu Köln, 5000 Köln 41, Federal Republic of Germany

*Communicated by Sydney Brenner, August 9, 1982*

**ABSTRACT** **A 10-kilobase DNA fragment containing exons 1 and 2 of the chicken lysozyme gene has been randomly cleaved with DNase I. After tailing and cloning into the plasmid pUK230, Lac$^+$ colonies were selected. Colonies harboring expressed fragments of the exons could be detected by an immunoenzymatic assay using antibodies against lysozyme. The smallest fragment coded for 10 amino acids and the largest coded for almost all residues of exon 2. These results suggest that any gene of any genome cloned in this way can be detected if antibodies against the gene product are available.**

Most eukaryotic genes are encoded not as contiguous stretches of DNA but rather as fragments, called "exons," interspersed with noncoding regions, called "introns" (1). Cloning of such DNA does not result in authentic gene products in *Escherichia coli*. The identification of such clones is only possible by physical methods such as colony or plaque hybridization, for which specific nucleic acid probes must be isolated (2–4). For genes that are expressed at low level the isolation of cDNA clones to be used as probes is time consuming or even not practicable. Methods have been described to overcome this problem. The most commonly used utilizes synthetic oligonucleotides as hybridization probes, which requires knowledge of the amino acid sequence of at least part of the gene product (5).

Recently we developed a method that allows the identification of gene fragments [down to 80 base pairs (bp)] from *E. coli* by the product such a fragment encodes (6). The gene fragments are cloned so that when they are expressed in *E. coli* their protein products are fused as $NH_2$-terminal peptides to β-galactosidase, thus stabilizing these protein products. The identification of colonies containing such chimeras was done by a variation of the antibody test used by Broome and Gilbert (7). The average length of eukaryotic exons is 100–200 bp. Their identification with the immunoenzymatic method thus should be possible after cloning of genomic DNA fragments 100–200 bp long. As a consequence, any eukaryotic gene could be directly cloned and identified if antibodies against the gene product were available.

To examine the possibilities of the method, we used a genomic DNA fragment from the chicken containing exon 1 and 2 of the lysozyme gene, the structure of which is well established (8).

## MATERIALS AND METHODS

**Strains and Plasmids.** *E. coli* K-12 BMH 71-18 (Δ[*lac pro*] *thi sulI F'lac I$^Q$Z* M15 *pro*$^+$) (9), pUR2 (10), and pUK230 (6) have been described, as have pls-1 (11) and λlys31 (12).

**Enzymes.** Restriction endonucleases *Bam*HI (New England BioLabs) and *Eco*RI and *Pst* I (Boehringer Mannheim) were used as described (13). DNA polymerase I (large fragment) was purchased from Bethesda Research Laboratories and terminal deoxyribonucleotidyltransferase was from P-L Biochemicals. The tailing reactions were done as described in ref. 14.

**Antibodies.** Rabbit antibodies against lysozyme were prepared with 0.1% $NaDodSO_4$-denatured electrophoretically purified lysozyme as the immunogen (Merck, Darmstadt, Federal Republic of Germany). They were purified from the sera as described (15).

**Preparation of Genomic DNA Fragments with DNase I.** The 7.0-kb *Eco*RI DNA fragment of λlys31 (12), containing exons 1 and 2 of the chicken lysozyme gene, was subcloned in the plasmid pUR2 (10). The subclone was called "pUR2-E7.0" (Fig. 1). pUR2-E7.0 (30 μg) was incubated in 100 μl of 33 mM Tris·HCl, pH 7.6/10 mM $MnCl_2$ containing 3 ng of DNase I (Boehringer Mannheim) at room temperature. After 10, 20, and 30 min, 33 μl of the reaction mixture was transferred to an Eppendorf tube containing 5 μl of 0.1 M EDTA at pH 8. The digested DNA was separated on a 6% polyacrylamide gel. DNA in the size range 100–200 bp was cut out and eluted overnight in TNE buffer (10 mM Tris·HCl, pH 7.6/150 mM NaCl/1 mM EDTA). The DNA eluate was applied to a DEAE-cellulose column (200 μl bed volume). The column was washed with 1 ml of TNE buffer, and the DNA was eluted with 200 μl of TNE buffer containing 1.5 M NaCl. DNA was precipitated with ethanol and tailed with terminal transferase and dC (14).

**Identification of Clones with the Immunoenzymatic Method.** Approximately 60 ng of dC-tailed genomic DNA fragments between 100 and 200 bp long and 200 ng DNA of the dG-tailed vector pUK230 were annealed in 10 μl of buffer (100 mM NaCl/10 mM EDTA, pH 8). Then, 3 μl of the annealing mixture was added to 0.2 ml of competent *E. coli* strain BMH 71-18 (9) at 0°C. After 10 min the cells were transferred to 42°C for 3 min. Then the cells were incubated with 1.5 ml of rich medium for 1 hr at 33°C, plated on ampicillin plates, and incubated overnight at 37°C. The transformation frequency was $10^4$ transformants per μg of vector DNA. Colonies were transferred by replica plating onto two lactose minimal plates and again incubated overnight. One plate, in a glass Petri dish, was used for the antibody screening (6). Positive clones were picked and tested again with antibodies against lysozyme (Fig. 2).

**Preparation of Plasmid DNA for Sequence Analysis.** A 5-ml overnight culture of plasmid-carrying cells was collected by centrifugation and lysed at room temperature with 1 ml of lysis mix (1% Brij 58/0.4% Na deoxycholate/50 mM Tris·HCl/50 mM EDTA, pH 7.6) after 10-min incubation with a mixture of 1 ml of 25% sucrose (in 50 mM Tris·HCl at pH 7.6), 0.1 ml of lysozyme solution (10 mg/ml), and 0.1 ml of 0.5 M EDTA. After centrifugation for 30 min at 20,000 rpm (Sorvall, SS34), the supernatant was added to 0.5 ml of a solution containing 5 M

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.

Abbreviation: bp, base pair(s).

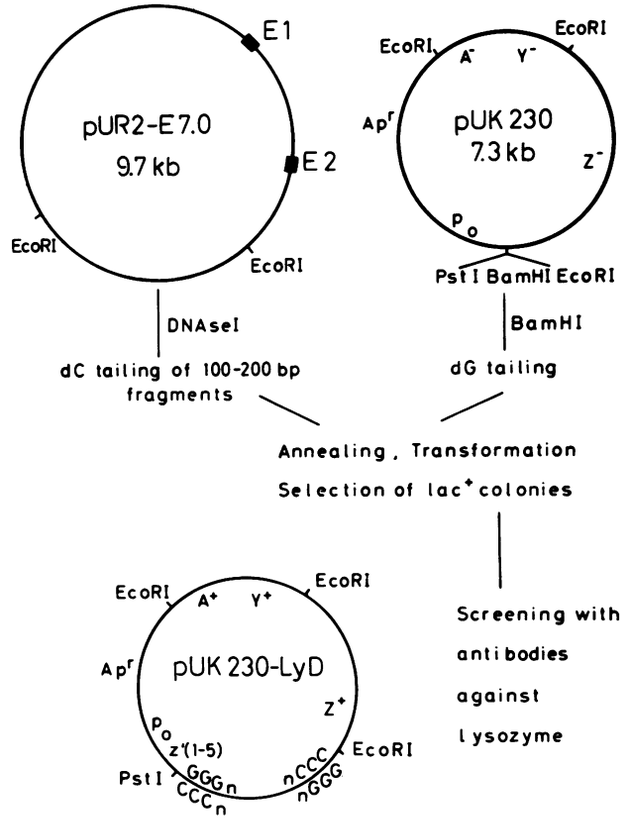


FIG. 1. General scheme of the experiments. A 7-kb *Eco*RI fragment of λlys31 (12) was cloned into the plasmid pUR2. The resultant plasmid pUR2-E7.0 contains exons 1 and 2 (E1 and E2) of the lysozyme gene. The vector pUK230 carries a frameshift mutation in the polylinker region (codons 5 and 6 of the *lac Z* gene). This reduces the background of plasmids without insert after tailing and annealing with the dC-tailed fragments. The plasmids of $Lac^+$ colonies must carry an insert as demonstrated by the plasmid pUK230-LyD. Thus, normally the chimeric protein begins with amino acid residues 1–5 of β-galactosidase followed by a poly (Gly) stretch $(GGG)_n$, followed by the residues encoded by the inserted DNA fragments, followed by a poly (Pro) stretch $(CCC)_n$ and residues 6–1,021 of β-galactosidase.

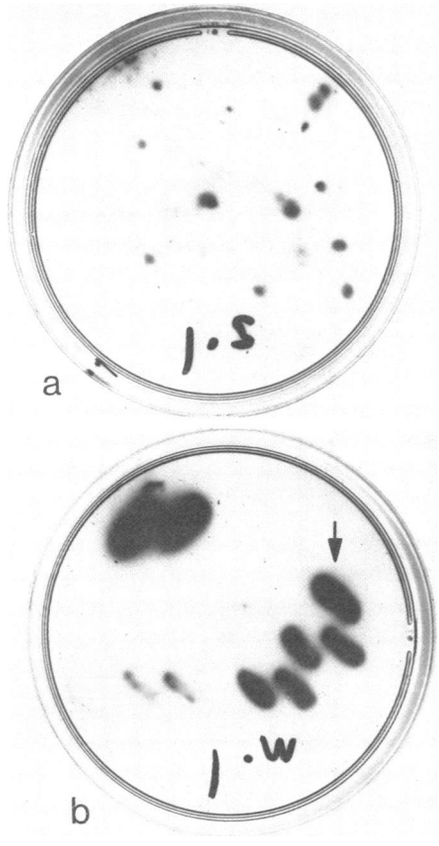


FIG. 2. Immunoenzymatic screening for colonies containing lysozyme specific sequences. (*a*) About 500 $Lac^+$ clones were tested on one plate for the presence of lysozyme–β-galactosidase chimeras. Such chimeras bound to antibodies against lysozyme fixed to a plastic sheet. They were visualized by hydrolysis of 5-bromo-4-chloro-3-indolyl β-galactoside (for details, see ref. 6). (*b*) Colonies corresponding to the blue marks were picked from the master plate, streaked twice on one plate and tested again. Approximately 50% of the clones picked gave a reproducible positive response and were subjected to sequence analysis. The arrow points to a control, a chimera between residues 8–84 of lysozyme and β-galactosidase (see text).

$NaClO_4$ (final concentration, 1 M). The solution was extracted twice with chloroform, and the DNA was collected by the addition of 0.5 vol of isopropanol and centrifugation for 5 min at 15,000 rpm (27,000 × *g*). The dried pellet was dissolved in 100 μl of DNA buffer (10 mM Tris·HCl/1 mM EDTA, pH 7.6); 10 μl of this DNA solution (0.5–1 μg) was digested with *Pst* I and *Eco*RI. Labeling, separation, and elution of the fragment was carried out as described (13). DNA sequence analysis was as described in ref. 16.

## RESULTS

**Construction and Identification of a Clone Containing Lysozyme cDNA.** The general scheme of the construction of the clones is shown in Fig. 1. The vector pUK230 (6) was cut with *Bam*HI in the promoter proximal region of the *lac Z* gene. Insertions in this region do not abolish activity of β-galactosidase (17). The dC-tailed fragment was annealed with the dG-tailed vector. A $Lac^-$ *E. coli* strain was transformed and $Lac^+$ bacteria were selected for. If the fragment contained no stop codon in phase and if it was inserted in the frame of β-galactosidase, $Lac^+$ clones should arise. After lysis of colonies, a plastic sheet coated with antibodies—in our case, against lysozyme—was used to detect positive colonies. Because only those β-galactosidase molecules that contain covalently bound antigen adhere to the antibody-coated plastic, the position of such clones can be readily visualized. A colorless substrate of β-galactosidase can be hydrolyzed to an insoluble colored (in our case blue) product.

For a first test of the antigenicity of a product made in *E. coli*, we constructed a clone that contained the major part of the chicken lysozyme mRNA sequence. We cloned the *Alu* I DNA fragment of a cDNA clone coding for amino acid residues 8–84 of the lysozyme (11) after dC tailing in the dG-tailed *Bam*HI site of vector pUK230. The DNA was transformed after annealing in the *E. coli* strain BMH 71-18 (a $Lac^-$ strain) (9). $Lac^+$ colonies were tested for the presence of expressed parts of the lysozyme gene with lysozyme-specific antibodies. Details of the test are described in ref. 6. Plasmid DNA from positive colonies was isolated and the sequences of the inserts were determined. All clones harbored the expected *Alu* I DNA fragment in the correct reading frame. We used one of these clones as a positive control in the determination of the antigenicity of the genomic clones.

**Construction and Identification of Clones Containing Fragments of Exon 1 or 2 of the Chicken Lysozyme Gene.** The general scheme of the construction was the same as above (Fig. 1). The positions in the genomic DNA sequence of the fragments that we found to give rise to antigenically active products are presented in Fig. 3, together with a partial sequence of the region of the part of the lysozyme gene we used in these exper-

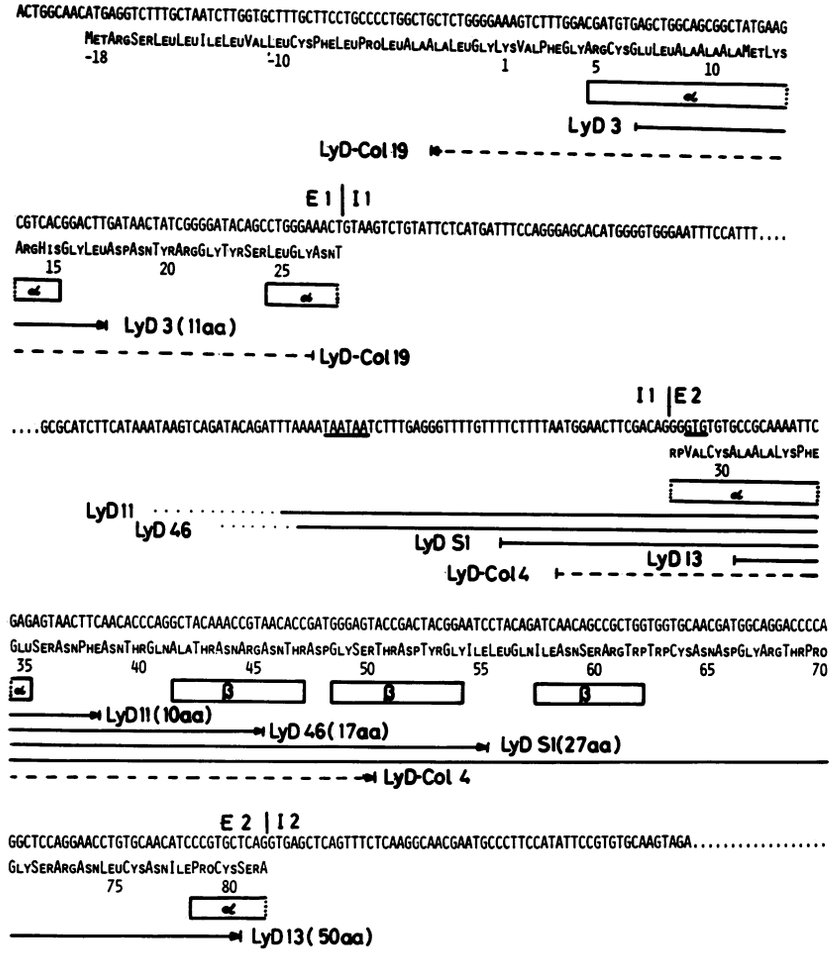


FIG. 3. Fragments of genomic lysozyme DNA detected with antibodies or by colony hybridization, or both. The DNA and protein sequences of part of the lysozyme gene, exons 1 and 2, are presented, along with the secondary structure of mature lysozyme ($\alpha$, $\alpha$-helical turn; $\beta$, $\beta$-sheets) (18). The lengths of the cloned fragments are shown with straight lines (detected with antibodies and colony hybridization) or with dashed lines (detected only by colony hybridization). Arrows show the orientation of the fragment cloned. The number of amino acids encoded by each fragment is indicated in brackets. Exons and introns are abbreviated E and I. The underlined bases T-A-A-T-A-A and G-T-G are the stop and start codons of clones LyD 11 and 46.

iments. We found fragments of both exons. The smallest exon fragment isolated contained only 10 codons (clone LyD 11) but part of intron 1 and two ochre stop codons in frame. This clone and also clone LyD 46 must start translations at the GUG of exon 2, corresponding to amino acid 29 of mature lysozyme. In the clone LyD 13, the exon DNA fragment coded for 50 amino acids. Here the complete exon 2 lacking the last and the first two codons was cloned.

The size of the DNA fragments used as starting material for cloning varied between 100 and 200 bp. The fragments recovered varied between 130 and 150 bp but one fragment, in clone LyD 3, was found to be only 33 bp long (Fig. 4). The antigenic strength of all these positive clones, measured as intensity of the blue color on the plastic sheets, was comparable to that of the cDNA clone and thus independent of the size of the fragment.

**Analysis of Lac$^+$ Clones by Colony Hybridization.** We reasoned that not all expressed exon DNA fragments might be detected by the immunoenzymatic screening method because not all fragments of this region capable of expression might carry an antigenic determinant. To get a rough idea of the frequency of such clones we used the method of colony hybridization (2). As a hybridization probe we used the lysozyme cDNA insert of plasmid pls-1 (11). We tested the 588 Lac$^+$ clones screened previously and found 4 additional clones containing exon DNA (Table 1). We subjected three of these clones to sequence analysis. Two were identical; they carried a DNA fragment of exon 1 in the opposite orientation but no stop codon in the $\beta$-galactosidase frame. The other clone contained a DNA fragment of exon 2 in the correct orientation but not in the frame of the lysozyme gene (Fig. 3). The lengths of these DNA fragments were 87 and 82 bp.

## DISCUSSION

Our results suggest the possibility of cloning and identifying directly fragments of exons from a genomic library without a nucleic acid probe. In the immunoenzymatic test 6 clones from 6,758 transformants were shown to contain parts of the lysozyme gene (Table 1). Thus, we found 1 specific clone among almost $10^3$ transformants. We did the experiment with a $10^4$-bp fragment. The genome of the malaria parasite *Plasmodium falciparum* contains about $10^7$ bp. Thus, it would be necessary to screen at least $10^6$ transformants ($10^3 \times 10^3$) if one wanted to detect a particular DNA sequence in this eukaryotic genome. But less than $10^5$ of these $10^6$ clones would be Lac$^+$ and only these have to be tested with the appropriate antibodies.

In our experiment we found about 10% of the clones to be Lac$^+$. If we were to perform the experiment with higher eu-

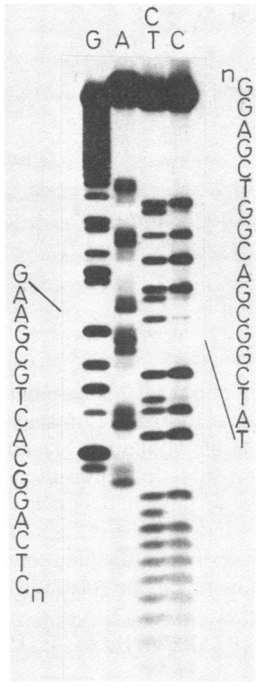


FIG. 4. Sequence of the insert of clone LyD 3 detected with antibodies against lysozyme. The plasmid DNA was isolated, and the inserted DNA was subjected to sequence analysis on a 12% polyacrylamide gel.

karyotic DNA (mouse, man), the proportion of $Lac^+$ clones probably would be much lower because the majority of such DNA does not code for any protein and is thus full of nonsense barriers. In fact, the separation of exons by introns brings an additional advantage to the method when one screens fragments 200–300 bp long because exons correspond in general to domains—i.e., folding units of proteins.

With the immunoenzymatic method we were able to identify 1 specific clone among $10^5$ $Lac^+$ colonies per Petri dish (data not shown). We conclude therefore from our model experiment that, in principle, genomic libraries could be screened with less than 100 Petri dishes. However, it may be advisable to screen a cDNA library constructed in the same way. After sequence analysis of the DNA fragments inserted in such clones, the frame of the DNA sequence in each case has been established because it is the same as the frame of β-galactosidase. An advantage is also presumably the stability of such a library. Cosmid libraries, for example, are relatively unstable because the large size of the inserts favors deletions. In our type of library the inserted DNA fragments are small (100–200 bp) and therefore deletions are less likely to occur during bacterial amplification.

Table 1. Frequency of clones containing exon structures

| Transformants, no. | $Lac^+$ transformants, no. | Clones detected, no. | |
|---|---|---|---|
| | | Antibody method | cDNA probe method |
| 4,420 | 588 | 4 | 8 |
| 2,338 | 311 | 2 | * |

* Not done.

Previous analysis of lysozyme suggested that at least some of its antigenic sites were composed of amino acid residues from different parts of the protein but not from a contiguous sequence (19). Cloning and expression of fragments of the lysozyme gene would not regenerate such structures and thus detection of these determinants with antibodies should be impossible. Because in our experiment with genomic lysozyme DNA we could demonstrate the opposite, this must mean that not all antigenic determinants of lysozyme are of this type. This is in line with some recent results (20).

One clone showed that 30 bp of exon DNA suffice for identification with antibodies. In one case the known three-dimensional structure of lysozyme suggests that the region is in the conformation of an α-helix; in the other case this is likely (Fig. 3).

The coding sequences for every individual contiguous antigenic site recognized by a series of monoclonal antibodies thus may be determined by this method. This is of interest because monoclonal antibodies that affect protein function could be used to identify the subdomain of the protein involved. The method may be also used for the production of vaccines. Small antigenic sites could be determined and then synthesized. After coupling to an insert carrier they could be used directly for vaccination. Finally, unknown targets in autoimmune diseases could be determined.

This work was supported by the Deutsche Forschungsgemeinschaft (Schwerpunkt Gentechnologie).

1. Gilbert, W. (1978) *Nature (London)* **271,** 501.
2. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 3961–3965.
3. Benton, W. D. & Davies, R. W. (1977) *Science* **196,** 180–182.
4. Williams, J. G. (1981) in *Genetic Engineering,* ed. Williamson, R. (Academic, London), Vol. 1, pp. 1–55.
5. Suggs, S. V., Wallace, R. B., Hirose, T., Kawashima, E. H. & Itakura, K. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 6613–6617.
6. Koenen, M., Rüther, U. & Müller-Hill, B. (1982) *EMBO J.* **1,** 509–512.
7. Broome, S. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 2746–2749.
8. Jung, A., Sippel, A. E., Grez, M. & Schütz, G. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5759–5763.
9. Messing, J., Gronenborn, B., Müller-Hill, B. & Hofschneider, H.-P. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 3642–3646.
10. Rüther, U. (1980) *Mol. Gen. Genet.* **178,** 475–477.
11. Sippel, A. E., Land, H., Lindenmaier, W., Nguyen-Huu, M. C., Wurtz, T., Timmis, K. N., Giesecke, K. & Schütz, G. (1978) *Nucleic Acids Res.* **5,** 3275–3294.
12. Lindenmaier, W., Nguyen-Huu, M. C., Lurz, R., Stratmann, M., Blin, N., Wurtz, T., Hauser, H. J., Sippel, A. E. & Schütz, G. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 6196–6200.
13. Rüther, U., Koenen, M., Otto, K. & Müller-Hill, B. (1981) *Nucleic Acids Res.* **9,** 4087–4098.
14. Roychoudhury, R. & Wu, R. (1980) *Methods Enzymol.* **65,** 43–62.
15. Groner, B., Hynes, N. E., Sippel, A. E., Jeep, S., Nguyen-Huu, M. C. & Schütz, G. (1977) *J. Biol. Chem.* **252,** 6666–6674.
16. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
17. Müller-Hill, B. & Kania, J. (1974) *Nature (London)* **249,** 561–563.
18. Imoto, T., Johnson, L. N., North, A. C. T., Phillips, D. C. & Rupley, J. A. (1972) in *The Enzymes,* ed. Boyer, P. D. (Academic, New York), Vol. 7, pp. 665–868.
19. Atassi, M. Z. & Lee, C. (1978) *Biochem. J.* **171,** 429–434.
20. Takagaki, Y., Hirayama, A., Fujio, H. & Amano, T. (1982) *Arch. Biochem. Biophys.* **214,** 730–762.