Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 10788–10791, December 1991
Biochemistry

Phasing the conformational unit of spectrin

(coiled coil/fodrin/dystrophin/α-actinin)

ENRIQUE WINOGRAD, DANIEL HUME, AND DANIEL BRANTON*

Department of Cellular and Developmental Biology, The Biological Laboratories, Harvard University, Cambridge, MA 02138

Contributed by Daniel Branton, August 19, 1991

ABSTRACT **Many proteins contain a repetitive sequence motif, which implies that they contain a repetitive structural motif. Spectrin and the related proteins dystrophin and α-actinin consist largely of repeated motifs of 100–120 residues. But the repeating motif is degenerate and it has been difficult to define the boundaries of the repeating sequence unit or its corresponding structural unit. We have determined at which residues the structural units that correspond to spectrin's repeating 106-amino acid motifs begin and end. *Drosophila* α-spectrin cDNAs were expressed in bacteria to show that single segments (106 amino acids) and pairs of segments encoded by selected regions of spectrin cDNA can fold into stable conformations whose biophysical and biochemical properties are similar to those of native spectrin. Because such folding was critically dependent on the phasing of the expressed sequence with respect to the apparent boundaries of the repeating motifs, our data provide experimental evidence that relates the boundaries of the folded, conformational unit to the chemical sequence of repeating motifs.**

Spectrins, some of which are also called fodrin, are elongated, flexible molecules whose primary structure is dominated by tandem, homologous, 106-residue motifs (1) that have been referred to as repeating units or segments (2). Dystrophin and α-actinin, other members of the spectrin superfamily (3, 4), contain comparable repetitive sequences. Although secondary-structure predictions suggest that each of the repetitive motifs of spectrin folds as a closely packed, triple α-helical unit, neither spectrin itself, nor proteolytically derived domains of spectrin, have been tractable samples for direct structure determination. Because the 106-residue motifs are generally repeated without interruptions, the residue that corresponds to the beginning (or end) of a structural unit is not self-evident. Consider, for example, the repetitive sequences of letters ABCDABCDABCD . . . . To produce a 4-letter segment of this sequence that can form a structural unit, it is essential to know whether the structure is ABCD, BCDA, CDAB, or DABC. Similarly, to express a 106-residue segment of spectrin that will fold into a structural unit, one must know at which position in any of the repetitive motifs that are evident in the primary sequence the segment should start and end. Speicher and Marchesi (1) stated clearly that the convention they adopted to illustrate the repetitive sequences was not intended to correlate with the conformational unit of the folded protein.

Using cDNA constructs derived from *Drosophila* α-spectrin and expressed in bacteria, we show that single repetitive motifs (106 amino acids) and pairs of motifs can fold into stable conformations similar to that of native spectrin when their N-terminal ends are 26 residues downstream of the residue that, according to the convention of Speicher and Marchesi (1), has usually been designated to be in the first position of each repetitive motif. Similarly sized polypeptides with other phasing do not fold into stable structures.

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.

MATERIALS AND METHODS

Sequence and Phasing of Constructs. cDNA clone 9AB (5), which encodes segments 10–22 of *Drosophila* α-spectrin, was used to generate all constructs. Constructs CLA and EAG were generated by digesting clone 9AB with *Cla* I or *Eag* I (Boehringer Mannheim), and the corresponding DNA fragments were purified and ligated to the pGEX-2T vector (6) by standard techniques (7). This vector expressed spectrin as a fusion protein linked, via a thrombin cleavage sequence, to glutathione transferase. For other constructs, selected regions of clone 9AB were amplified by the polymerase chain reaction (7). The oligonucleotide primers contained *Eco*RI sites that were used to facilitate cloning the amplified DNA into the corresponding site of pGEX-2T; those corresponding to the 3′ end of the amplified DNA included a translation stop codon to minimize the number of foreign residues in the expressed protein. All clones containing DNAs initially generated by the polymerase chain reaction were sequenced to verify faithful amplification.

Expression and Protein Purification. Spectrin was purified from human erythrocytes; recombinant fragments were expressed in and purified from *Escherichia coli* (strain 71-18). Overnight cultures were diluted 1:50 in fresh LB+ (7) with ampicillin (100 μg/ml) and cells were grown at 37°C for 90 min. Isopropyl β-D-thiogalactopyranoside was added to 0.5 mM final concentration and growth was continued for 2 hr. Cells were harvested and then were lysed by sonication. The expressed glutathione transferase–spectrin fusion proteins were affinity-purified on glutathione-agarose beads (Sigma), and the spectrin fragments were cleaved off with thrombin (6) and purified to homogeneity on a Mono Q column (Pharmacia FPLC system).

Circular Dichroism. Just prior to CD measurements, proteins were diluted to 0.05–0.4 mg/ml in 5 mM sodium phosphate buffer, pH 7.5/150 mM NaCl and centrifuged for 30 min at 40,000 × *g* to remove aggregated material. Protein was determined by a modified Kjeldahl technique (8) using nitrogen ratios from the predicted amino acid sequence. Circular Dichroism measurements were made in an Aviv model 60DS spectropolarimeter equipped with a thermostatted cell housing. CD spectra were recorded at 0.25-nm steps using a cell of 1-mm path length. The first Cotton effect was measured at 25°C using a cell of 1-cm path length. Denaturation experiments were carried out from 25°C to 75°C at 222 nm and the midpoint denaturation temperature was estimated by taking the first derivative.

Proteolytic Digestions and N-Terminal Sequencing. Proteins were dialyzed against 50 mM NaCl/20 mM Tris·HCl, pH 7.5. Incubation with α-chymotrypsin or elastase was at room temperature at the indicated enzyme/substrate ratio. Immediately after digestions, timed samples were withdrawn,

*To whom reprint requests should be addressed.

boiled in SDS/PAGE sample buffer, and electrophoresed in Tricine/SDS/10% polyacrylamide gels (9). Digests for N-terminal sequencing were electroblotted onto poly(vinylidene difluoride) membranes (Millipore) and sequenced on an Applied Biosystems model 470 sequenator equipped with an on-line 120A phenylthiohydantoin analyzer (Microsequencing Facility, Harvard University).

RESULTS

To verify that bacterially expressed *Drosophila* α-spectrin fragments are soluble polypeptides whose conformation would be representative of native spectrin, we expressed a 3.2-kilobase-pair *Eag* I restriction fragment of *Drosophila* α-spectrin. The encoded protein, EAG, included repetitive segments 11 through 19 and portions of segments 10 and 20 (residues 1020–2091 of *Drosophila* α-spectrin). EAG's protease resistance (Table 1) was similar to that of erythrocyte spectrin (10), its CD spectrum was characteristic of an α-helical protein (Fig. 1), and its molar ellipticity was comparable to that of native erythroid spectrin (Table 1). As previously shown for other fragments of spectrin (11), EAG folded into structures whose EM images resembled the corresponding regions of the native molecule (data not shown) and whose fractional contour length (48 ± 5% as long as authentic *Drosophila* α-spectrin; ref. 11) was not significantly different from its fractional sequence length (44% of the residues in full-length α-spectrin). These data indicate that α-spectrin expressed from its corresponding cDNA in bacteria can adopt a conformation that is similar to that of native spectrin.

Although we did not know at which position in the repetitive sequences each structural unit starts and ends, we reasoned that a polypeptide consisting of more than one but less than three spectrin motifs might fold as one complete, stably folded unit with unstably folded, or "floppy," protease-sensitive ends. To test this idea, a 687-base-pair *Cla* I cDNA restriction fragment that encodes 229 residues centered around segment 14 (Fig. 2) was expressed and purified. In contrast to spectrin or EAG, this 229-residue protein, CLA, exhibited a CD spectrum with a lower minimum at 208 nm than at 222 nm (Fig. 1). Since the CD spectrum of polypeptides in the random coil conformation present a minimum at ≈200 nm, the CD spectrum of CLA was consistent with the notion that this protein contained a well-folded α-helical domain with floppy or random coil ends.

Table 1. Properties of spectrin and its fragments

Protein	Size, amino acids	$[\theta]_{222}$*	Protease product(s)†
Spectrin‡	4566	27,375	25,000 (T)
EAG	1071	27,131	29,000 (C or E)
CLA	229	16,002	14,500 (C)
CCP	122	25,897	14,500§ (C)
14B	108	(8,839)¶	<3,000
B14	109	26,827	12,800§ (E)
B15	109	ND	12,800§ (E)
B14-15	215	ND	25,600§ and 12,800 (E)

*Molar ellipticity values at 222 nm ($[\theta]_{222}$, degree·cm^2·dmol^{-1}) were based on at least three independent CD and protein determinations; SD was ±5%; ND, not determined.

†The apparent molecular weight, based on SDS/PAGE, of the *smallest* chymotrypsin (C), elastase (E), or trypsin (T, from ref. 10)-resistant polypeptide is indicated.

‡Human erythrocyte spectrin was purified by conventional methods.

§Same as before proteolysis.

¶The determined value cannot be attributed to polypeptide conformation, as it exhibited no melting behavior.

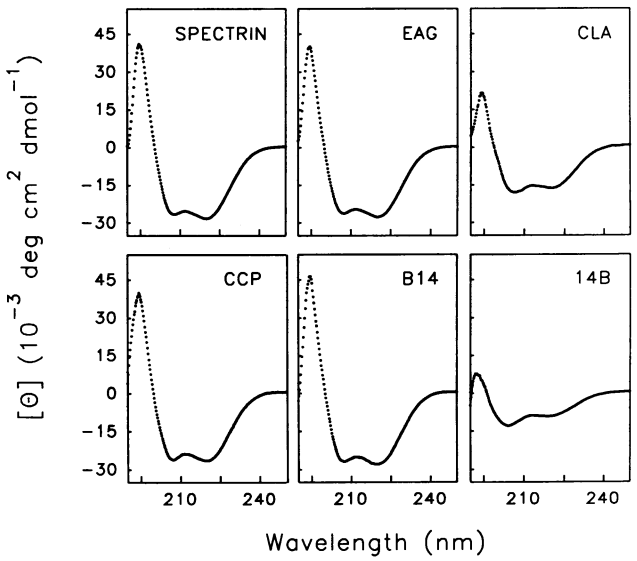

FIG. 1. CD spectra of indicated spectrin fragments.

To test this idea, CLA was incubated with either α-chymotrypsin or elastase (Fig. 3*a*). Although the two enzymes have different substrate specificities, the pattern of digestion was similar: both proteases produced $M_r \approx 14{,}500$ polypeptides that were resistant to further proteolysis. Furthermore, N-terminal sequencing showed that, in producing the M_r 14,500 polypeptide, chymotrypsin had cleaved CLA down to Threonine-1383, within 11 residues of where elastase had cleaved CLA, at alanine-1372. This digestion pattern showed that CLA contained a single protease-resistant core with protease-sensitive ends.

To assess the properties of the protease-resistant core, we designed and expressed a protein equivalent to the smallest protease-resistant fragment generated by chymotrypsin digestion. The N terminus of the protein was selected as threonine-1383 (residue 20 in segment 14). The C terminus was then deduced by inspecting the *Drosophila* α-spectrin sequence to identify an appropriate downstream chymotrypsin cleavage site that would produce a polypeptide whose molecular weight would be about 14,500. Phenylalanine-1504 (residue 35 in the 15th motif) was selected. The corresponding region of cDNA was amplified by the polymerase chain reaction, and subcloned into pGEX-2T, and the expressed protein, CCP (Fig. 2), was purified. As a control, P1, a polypeptide with the same number of residues as CCP but that starts at a different position in segment 14 (Fig. 2), was similarly generated.

As expected, CCP migrated as a $M_r \approx 14{,}500$ protein in SDS/polyacrylamide gels and was completely resistant to chymotrypsin digestion. In contrast, P1 was rapidly degraded (Fig. 3*b*). CD showed that CCP was predominantly α-helical: its 208 nm/222 nm molar ellipticity ratio was smaller than that of CLA (Fig. 1) and its molar ellipticity value at 222 nm (Table 1) was substantially larger than that of CLA, as expected of a protein that lacked the protease-sensitive sequences that we had postulated to account for CLA's greater ellipticity at 208 nm than at 222 nm.

Because CCP extends from residue 20 in segment 14 to residue 34 in segment 15, it contains 15 more residues than a single 106-amino acid motif. Therefore, four smaller proteins, designated B14B, TB14, 14B, and B14 (Fig. 2), were designed and tested by incubation with elastase (Fig. 3*d*). Of the four, only B14 ($M_r \approx 12{,}800$), whose N and C termini were both located within one residue of position 27 in two neighboring motifs, appeared to include all of its residues in a stably folded, elastase-resistant structure. The others were rapidly

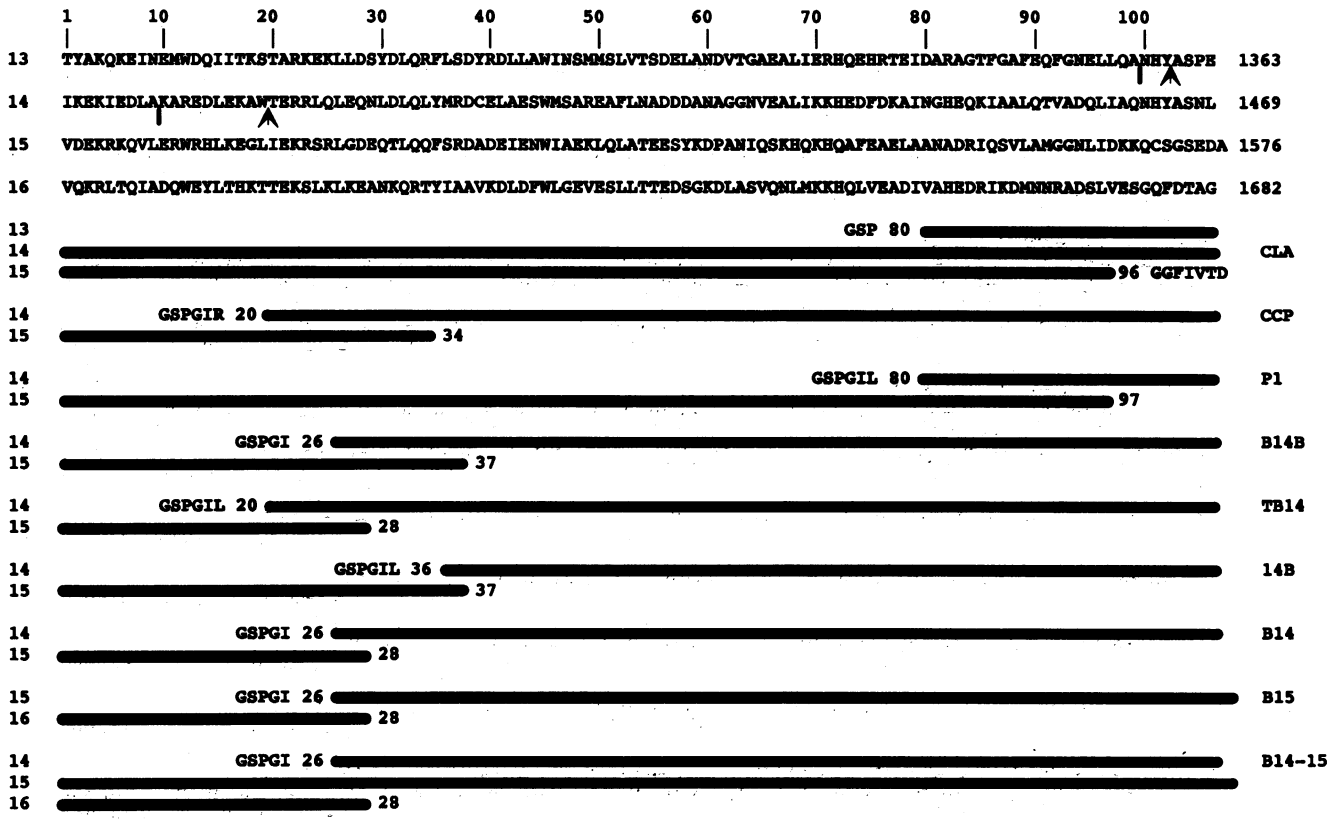

FIG. 2. Sequence and phasing of constructs. The first four lines show *Drosophila* α-spectrin residues 1258–1682 in segments 13–16 displayed according to the convention of Speicher and Marchesi (1). The cleavage sites that produce the M_r 17,000 and 14,500 polypeptides in CLA (Fig. 3) with elastase (vertical bars) or chymotrypsin (arrows) are indicated. These sites were determined from the N-terminal sequence of the corresponding fragments. Numbers at right of the first four lines indicate the residue position of the last amino acid in each line; those at left indicate the segment number according to the convention of Speicher and Marchesi (1). The constructs used in this study are aligned below the sequences, with solid lines indicating the included residues from the sequences above. Any non-spectrin residues, a by-product of the expression system, are shown in single-letter abbreviation beyond the construct termini.

proteolyzed to yield either a single protease-resistant fragment that in SDS/PAGE behaved exactly as B14 (B14B and TB14) or small polypeptides that could not be resolved by SDS/PAGE (14B). CD measurements (Fig. 1 and Table 1) indicated that B14 was largely α-helical with molar ellipticity values at 222 nm that were similar to those of native spectrin.

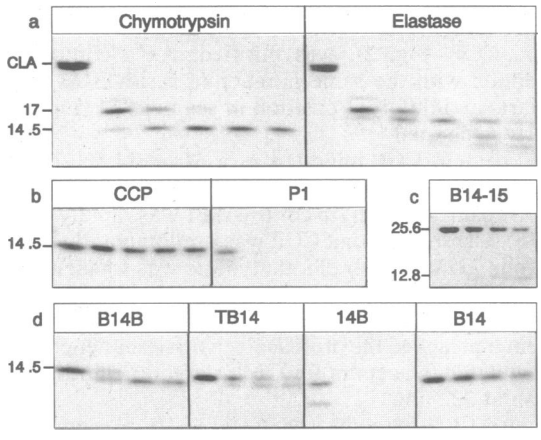

FIG. 3. Spectrin fragment susceptibility to protease. Numbers at left indicate $M_r \times 10^{-3}$. (*a*) Successive lanes in each panel show CLA after incubation with enzyme (enzyme/substrate ratio = 1:160) for 0, 15, 30, 60, 90, or 120 min. (*b*) CCP or P1 was incubated with chymotrypsin (1:400) for 0, 15, 30, 60, 90, or 120 min. (*c*) B14-15 was incubated with elastase (1:100) for 0, 15, 30, or 90 min. (*d*) Indicated fragments were incubated with elastase (1:100) for 0, 30, 60, or 90 min.

The midpoint denaturation temperature of B14 was 53°C. Together, the protease resistance and CD data show that the boundaries of B14 define a well-folded polypeptide representative of spectrin's repeating structural units whereas the other proteins did not. These results also provide direct experimental evidence supporting the suggestion (11) that interactions within each repetitive motif are sufficient to direct its proper folding.

Several other polypeptides were expressed to confirm the generality of the segment boundary positions defined by B14, including one, B15, whose boundary positions correspond to B14 but are shifted downstream 106 amino acids, and another, B14-15, whose boundary positions are similar to those of B14 but include two complete motifs (Fig. 2). As expected, B15 was completely resistant to chymotrypsin and elastase (data not shown). Although B14-15 was resistant to chymotrypsin (data not shown), elastase did slowly cleave B14-15 to a resistant $M_r \approx 12{,}800$ protein (Fig. 3*c*), consistent with cleavage at or near the middle of this two-segment construct. The resistance of B14-15 to chymotrypsin and its slow cleavage by elastase indicate that accessibility to any protease-recognition sites between the two conformational units must be restricted by the structure of the two contiguous segments. Access to potential cleavage sites between repeating segments is similarly restricted in intact spectrin (10) and in EAG (Table 1).

DISCUSSION

Since the initial hypothesis regarding spectrin structure was advanced (1), more extensive analysis and additional se-

quence information about spectrin and the related protein dystrophin have led to variations on the original model. Some of these show a different phasing of the conformational units (2, 12), others postulate more β structure (13, 14) or a longer α-helix rather than a flexible loop between structural units (5, 15–17), and several emphasize the possibility of coiled-coil interactions between the helices within each segment (15–17). Parry and Cohen (17) explicitly predicted an independently folded coiled-coil unit nearly identical in phasing to B14. While we have not expressed polypeptides that sample all possible phasings, the position of the protease-resistant portion within CLA together with the CD spectra and protease susceptibility of differently phased polypeptides, such as 14B and others that we have examined (data not shown), are consistent only with the folding of structured units whose phasing is similar to B14 or B15. Furthermore, the low molar ellipticity at 222 nm of CLA (Table 1) and the protease susceptibility (Fig. 3*d*) of the partial segments that extend beyond the one complete segment contained within CLA, B14B, and TB14 indicate that, if the connection between segments is α-helical in native spectrin, both this α-helical structure and its protease resistance are dependent on intrasegment interactions, presumably coiled-coil interactions (18).

Although we have examined *Drosophila* α-spectrin and probed only segments 14 and 15, the homologous repeating segments that have been derived from a common ancestral gene in all of the spectrin superfamily (3, 4) should have similar conformations, since chain folding is even more strongly conserved than amino acid sequence (19). Our results should therefore apply to all of the repeating segments in all of the spectrins and probably apply also to other members of the spectrin superfamily, including dystrophin (12) and α-actinin.

In fact, our evidence showing where the conformational unit begins and ends implies a display and alignment of the chemical sequences in the spectrin superfamily that bring into register a larger number of their consensus sequences than when they are displayed according to the convention of Speicher and Marchesi (1). Furthermore, our data suggest that the first complete structural unit (the first complete segment) of spectrin begins very near the consensus leucine residue that aligns with position 26 of Fig. 2 (at approximately leucine-43 of *Drosophila* α-spectrin). Thus, α-spectrin appears to begin with a partial segment. Tse *et al.*'s (20) recent studies of mutant spectrin forms bear on this conclusion. They accounted for altered dimer self-association and protease susceptibility by proposing that a partial segment near the carboxyl end of β-spectrin associates with a partial segment at the amino end of α-spectrin to form a complete, closely packed, triple α-helical unit. Our data support their hypothesis by showing that there is a partial segment at the amino end of α-spectrin and by showing that partial segments are very susceptible to proteolysis.

The approach we have used to define the boundaries of spectrin's conformational units should be applicable to other proteins that contain repeating segments whose boundaries are not immediately evident in fragments derived by proteolytic digestion of the whole protein. Knowledge of the segment boundaries that favor the correct folding of expressed polypeptides in the spectrin family will facilitate efforts to devise assays that define the specific binding properties and biochemical functions of selected regions within these large multifunctional proteins.

We thank Tim Cronin for excellent technical assistance; Felix Rey, Peter S. Kim, Carolyn Cohen, and Ronald Dubreuil for helpful suggestions and discussions; and Terry Tao and Samuel Lehrer for allowing use of the spectropolarimeter. This work was supported by National Institutes of Health Grants HL17411 and GM39686 and by Damon Runyon–Walter Winchell Cancer Research Fund Fellowship DRG-998 (to E.W.).

1. Speicher, D. W. & Marchesi, V. T. (1984) *Nature (London)* **311,** 177–180.
2. Speicher, D. W. (1986) *J. Cell. Biochem.* **30,** 245–258.
3. Byers, T. J., Husain-Chishti, A., Dubreuil, R. R., Branton, D. & Goldstein, L. S. B. (1989) *J. Cell Biol.* **109,** 1633–1641.
4. Davison, M. D., Baron, M. D., Critchley, D. R. & Wootton, J. C. (1989) *Int. J. Biol. Macromol.* **11,** 81–90.
5. Dubreuil, R. R., Byers, T. J., Sillman, A. L., Bar-Zvi, D., Goldstein, L. S. B. & Branton, D. (1989) *J. Cell Biol.* **109,** 2197–2205.
6. Smith, D. B. & Johnson, K. S. (1988) *Gene* **67,** 31–40.
7. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
8. Jaenicke, L. (1974) *Anal. Biochem.* **61,** 623–627.
9. Schagger, H. & Jagow, G. V. (1987) *Anal. Biochem.* **166,** 368–379.
10. Speicher, D. W., Morrow, J. S., Knowles, W. J. & Marchesi, V. T. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5673–5677.
11. Dubreuil, R. R., Brandin, E., Reisberg, J. H. S., Goldstein, L. S. B. & Branton, D. (1991) *J. Biol. Chem.* **266,** 7189–7193.
12. Koenig, M. & Kunkel, L. M. (1990) *J. Biol. Chem.* **265,** 4560–4566.
13. Marchesi, S. L., Letsinger, J. T., Speicher, D. W., Marchesi, V. T., Agre, P., Hyun, B. & Gulati, G. (1987) *J. Clin. Invest.* **80,** 191–198.
14. Xu, Y., Prabhakaran, M., Johnson, M. E. & Fung, L. W. M. (1990) *J. Biomol. Struct. Dyn.* **8,** 55–62.
15. Koenig, M., Monaco, A. P. & Kunkel, L. M. (1988) *Cell* **53,** 219–228.
16. Cross, R. A., Stewart, M. & Kendrick-Jones, J. (1990) *FEBS Lett.* **262,** 87–92.
17. Parry, D. A. D. & Cohen, C. (1991) *AIP Conf. Proc.* **226,** 367–377.
18. Cohen, C. & Parry, D. A. D. (1990) *Proteins Struct. Func. Genet.* **7,** 1–15.
19. Pink, J. R. L. (1981) in *Biochemical Evolution*, ed. Gutfreund, H. (Cambridge Univ., New York), pp. 231–260.
20. Tse, W. T., Lecomte, M.-C., Costa, M., Garbarz, F. F., Feo, C., Boivin, P., Dhermy, D. & Forget, B. G. (1990) *J. Clin. Invest.* **86,** 909–916.