

# Molecular Cloning of cDNA for Caltractin, a Basal Body-associated Ca<sup>2+</sup>-binding Protein: Homology in Its Protein Sequence with Calmodulin and the Yeast *CDC31* Gene Product

Bessie Huang, Adele Mengersen, and Vincent D. Lee

Department of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, California 92037

**Abstract.** An extended synthetic oligonucleotide (59-mer) was used to isolate a *Chlamydomonas* cDNA containing the entire coding region for a novel basal body-associated 20-kD calcium-binding protein (CaBP). DNA and RNA blot analysis indicate that the 20-kD CaBP is encoded by a single copy gene from which is derived an ~1.1-kb-long transcript. The deduced amino acid sequence for the protein shows a linear relatedness with calmodulin from *Chlamydomonas* and other organisms (45–48% identity). The primary protein sequence of the 20-kD CaBP and its predicted secondary structure suggests that the protein is likely to contain four homologous calcium-binding

domains that conform to the helix-loop-helix (or EF hand) structure found in calmodulin and related calcium-modulated proteins. The major difference between the protein and calmodulin is an amino-terminal domain of 21 amino acids present on the 20-kD CaBP. In addition to its relatedness to calmodulin, the *Chlamydomonas* 20-kD CaBP shows a strong sequence identity (50%) with the yeast *Saccharomyces cerevisiae* *CDC31* gene product required for spindle pole body duplication. The association of these sequence-related calcium-binding proteins to microtubule-organizing centers of divergent structure suggests a potential conserved function for the proteins.

CALCIUM exerts many of its biological effects as a signal transducer in eukaryotic cells through its reversible interaction with a class of proteins referred to as calcium-modulated proteins (for reviews see references 18, 35). As a family, the calcium-modulated proteins, which include calmodulin, troponin C, parvalbumin, intestinal calcium-binding protein, and S100, are characterized by a low  $M_r$ , acidic pIs, and high affinity for calcium. Three of the calcium-modulated proteins (calmodulin, troponin C, and parvalbumin) have been found to exhibit a high degree of sequence homology (6, 36). Based on the structure of parvalbumin, Kretsinger and Barry (19) proposed a helix-loop-helix (or EF hand) conformation for the calcium-binding domains of calcium-modulated proteins. This proposed structure for the calcium-binding domains has been confirmed in the x-ray crystal structure determinations for the intestinal calcium-binding protein (32), troponin C (12, 31), and calmodulin (2, 20).

In the accompanying paper (13), we described the identification and purification of a novel 20-kD calcium-binding protein (CaBP)<sup>1</sup> isolated from the unicellular green alga *Chlamydomonas reinhardtii*. This protein, which was purified from isolated basal body complexes, the major microtubule organizing center in the cell, showed several biochemical

features in common with members of the calcium-modulated family of proteins. It is characterized by a low  $M_r$  of 20 kD, an experimentally determined pI of 5.3, an altered electrophoretic mobility in SDS-polyacrylamide gels in the presence of added calcium, and a calcium-dependent binding affinity to the hydrophobic ligand phenyl-Sepharose. The amino acid composition of the protein and partial peptide sequence analysis provided evidence that the 20-kD CaBP was clearly related to, but distinct from, other well-characterized members of the family of calcium-modulated proteins.

In this study, we describe the isolation and characterization of a cDNA that contains the entire coding sequence for the *Chlamydomonas* 20-kD CaBP. The deduced amino acid sequence of the protein shows a strong sequence relatedness with the experimentally determined sequence for *Chlamydomonas* calmodulin (22) and the deduced amino acid sequence of the yeast *CDC31* gene product required for spindle pole body duplication (3).

## Materials and Methods

### Oligonucleotide Synthesis and Labeling

DNA oligonucleotides were synthesized with a model No. 380B synthesizer (Applied Biosystems, Inc., Foster City, CA). For screening of the cDNA library, a 59-mer oligonucleotide corresponding to the amino acid sequence of a selected region of the *Chlamydomonas* 20-kD CaBP was synthesized and end-labeled with ( $\gamma$ -<sup>32</sup>P) ATP (Amersham Corp., Arlington Heights,

1. Abbreviation used in this paper: CaBP, calcium-binding protein.

IL) and T4 polynucleotide kinase (Pharmacia Fine Chemicals, Piscataway, NJ) as described by Maniatis et al. (23).

### Isolation of cDNA Clones

A cDNA library constructed in lambda gt10 from *Chlamydomonas* polyadenylated RNA (10) was kindly provided by Dr. Steve Mayfield (Research Institute of Scripps Clinic). Approximately 250,000 phage from this library were screened by hybridization with the radiolabeled 59-mer oligonucleotide probe at 50°C using the methods described by Whitehead et al. (37). Duplicate nitrocellulose (Schleicher & Schuell, Inc., Keene, NH) filters were washed using reduced stringency conditions in 6× SSC, 0.05% NaPPi at 25 or 37°C. Eight positive clones were plaque purified and the phage DNA was isolated using either DEAE-cellulose columns (11) or the phage adsorbent, LambdaSorb (Promega Biotek, Madison, WI). The Eco RI inserts were subcloned into pUC19 and M13mp19 (Pharmacia Fine Chemicals, Inc.) using standard techniques (23).

Restriction endonuclease mapping and partial DNA sequence analysis revealed that three of the eight clones contained overlapping, partial cDNAs for the 20-kD CaBP (pCaBP-1, -2, and -3). Of the remaining five clones, three contained overlapping cDNAs for calmodulin (based on DNA sequence analysis), and two contained inserts that did not hybridize with either the 20 kD CaBP or calmodulin cDNAs under high stringency hybridization and washing conditions (as described below).

A clone containing the entire coding region of the 20-kD CaBP was obtained by screening another aliquot of the original *Chlamydomonas* cDNA library with a 200-bp Eco RI/Xho I fragment from the 5' end of the cDNA insert in pCaBP-2. The cDNA fragment was isolated from low melt agarose and oligo-labeled as described below. Approximately 215,000 plaques transferred to replica nitrocellulose filters and were screened under high stringency conditions of hybridization and washing. Hybridization with the radiolabeled 5'-cDNA probe was carried out in 6× SSC, 1× Denhardt's, 0.1% SDS, 100 µg/ml salmon sperm DNA at 65°C for 48 h. The hybridized filters were initially washed at room temperature in 2× SSC and 0.1% SDS, then at 65°C in 0.2× SSC, 0.1% SDS, and finally at 65°C in 0.1× SSC, 0.1% SDS. Eight positive clones were plaque purified and the size of the cDNA inserts was determined. The phage containing the longest insert (pCaBP-4) was subcloned into pUC19 and M13mp19.

### DNA Sequence Analysis

Eco RI-digested DNA isolated from positive plaques was subcloned into M13mp19 using standard methods (23). DNA sequence data were obtained using the dideoxynucleotide method with both the Klenow fragment of DNA polymerase I (29) from Pharmacia Fine Chemicals, Inc. and modified T7 DNA polymerase (33) from United States Biochemical Corp. (Cleveland, OH). The 20-kD CaBP cDNA inserts were sequenced using the 17-mer universal primer (Pharmacia Fine Chemicals), the 59-mer synthetic oligonucleotide used in the initial screening of the cDNA library, and synthetic oligonucleotides prepared against previously sequenced regions as primers. Both strands of the cDNA insert in pCaBP-4 were completely sequenced. DNA sequences were compiled and analyzed using PC Gene software (Intelligenetics, Palo Alto, CA). Amino acid sequence comparisons and database searches were carried out with Microgenie software (Beckman Instruments, Inc., Palo Alto, CA) using the National Biomedical Research Foundation (NBRF) Protein Data Bank.

### Radiolabeling of cDNA

Eco RI- and Eco RI-Xho I-digested plasmid DNA containing the 20-kD CaBP cDNA was separated by electrophoresis in low-melting temperature agarose (IBI, New Haven, CT). The cDNA insert and restriction fragments derived from the inserts were labeled with <sup>32</sup>P-dCTP (Amersham Corp.) to high specific activity using random hexanucleotides (Amersham Corp.) as primers as described previously (7).

### Isolation and Blot Analysis of RNA

Total RNA was prepared from wild-type *Chlamydomonas* as previously described by Lefebvre et al. (21) with the following modification: RNA precipitates were washed twice with 2 M LiCl to remove polysaccharides (1) before they were further purified by centrifugation through a CsCl cushion (9). Poly(A) RNA was isolated by two cycles of oligo(dT)-cellulose chromatography (23). The RNA was separated on a 1.5% agarose/formaldehyde gel as described by Maniatis et al. (23) except that formaldehyde was

added to a final concentration of 0.66 M (5), transferred to a nylon filter (Magna Nylon 66; Micron Separations, Inc., Honeoye Falls, NY), and baked for 2 h at 80°C. Hybridization with cDNA probes, prepared and radiolabeled as described above, was performed at 42°C for 27 h in 50% formamide, 5× SSPE, 2× Denhardt's, and 250 µg/ml sonicated salmon sperm DNA. After hybridization, the filter was washed three times (10 min each) in 2× SSPE and 0.1% SDS at room temperature, three times (10 min each) in 0.5× SSPE and 0.1% SDS at 65°C, and finally three times (10 min each) in 0.2× SSPE and 0.1% SDS at 65°C before exposure to Kodak XAR-5 film for autoradiography at -75°C with an intensifying screen.

### Isolation and Blot Analysis of Genomic DNA

Genomic DNA was prepared from wild-type *Chlamydomonas* as previously described (26). DNA was digested with restriction enzymes, separated on a 0.7% agarose gel, and transferred to a nylon filter (Gene Screen; New England Nuclear, Boston, MA). After baking, the filter was hybridized with cDNA probes (prepared and radiolabeled as described above) at 42°C for 72 h in 50% formamide, 5× SSPE, 2× Denhardt's, and 250 µg/ml sonicated salmon sperm DNA. After hybridization, the filter was washed three times in 2× SSPE and 0.1% SDS at room temperature, and then three times in 0.5× SSPE and 0.1% SDS at 60°C before autoradiography. Each wash was for 30 min.

## Results

### Isolation of cDNA Clones for the Basal Body-associated Ca<sup>2+</sup>-binding Protein

A *Chlamydomonas* cDNA library was screened with a 59-base synthetic oligonucleotide probe (Fig. 1). The sequence of the oligonucleotide corresponded to the first 20 amino acids of a 31-residue internal linear sequence previously determined by microsequencing the 20-kD CaBP (13). In synthesizing the oligonucleotide, the choice of codons for the different amino acids was made according to the biased codon usage reported for the β-tubulin genes in *Chlamydomonas* (39). The oligonucleotide used as the probe was synthesized to be complementary to the coding sequence.

In the initial screening of 250,000 plaques of the *Chlamydomonas* cDNA library with the <sup>32</sup>P-labeled 59-mer, eight clones were identified that gave positive signals on duplicate filters washed under low stringency conditions (see Materials and Methods). The positive plaques were purified to homogeneity and the cDNA inserts subcloned into pUC19 and M13mp19. Restriction endonuclease mapping and partial DNA sequence analysis revealed that three of the eight clones (pCaBP-1, -2, and -3) contained overlapping, polyadenylated, partial cDNAs for the 20-kD CaBP (Fig. 2). Of the remaining five clones, three contained overlapping cDNAs for calmodulin, and two contained inserts distinct from the 20-kD CaBP and calmodulin cDNAs (data not shown).

To obtain a clone containing the entire coding region of the 20-kD CaBP, a 200-bp 5' Eco RI-Xho I fragment isolated from pCaBP-2 was labeled with <sup>32</sup>P-dCTP and was used to screen another aliquot of the same *Chlamydomonas* cDNA library. Approximately 215,000 plaques were screened under high stringency conditions (see Materials and Methods). Eight plaques that gave positive signals on replica filters were purified to homogeneity. The phage DNA was digested with Eco RI and Eco RI-Xho I, blotted, and hybridized with the <sup>32</sup>P-labeled 200-bp 5' probe. One of the clones, pCaBP-4, was characterized as containing the longest insert with an ~350-bp Eco RI-Xho I fragment that hybridized with the 5' probe. pCaBP-4 was subcloned into pUC19 and M13mp19 in

<u>Peptide</u>	Glu-Leu-Gly-Glu-Asn-Leu-Thr-Glu-Glu-Glu-Leu-Gln-Glu-Met-Ile-Ala-Glu-Ala-Asp-Arg
<u>Codons</u>	GAG CT <sub>C</sub> <sup>G</sup> GGC GAG AAC CT <sub>C</sub> <sup>G</sup> ACC GAG GAG GAG CT <sub>C</sub> <sup>G</sup> CAG GAG ATG ATC GCC GAG GCC GAC CG
<u>cDNA Sequence</u>	GAG CTG GGC GAG AAC CTG ACT <sup>*</sup> GAG GAG GAG CTC CAA <sup>*</sup> GAG ATG ATT <sup>*</sup> GCG <sup>*</sup> GAG GCG <sup>*</sup> GAC CG

Figure 1. The amino acid sequence and corresponding nucleotide sequence used to generate a synthetic oligonucleotide probe for the isolation of the *Chlamydomonas* 20-kD CaBP cDNA. The experimentally determined amino acid sequence of a selected region of the *Chlamydomonas* 20-kD CaBP (13) is shown. Underneath the amino acids are the codons chosen to generate a corresponding coding nucleotide sequence. The selection of codons was based on the biased codon usage reported for the *Chlamydomonas* nuclear-encoded  $\beta$ -tubulin genes (39). The 59-mer oligonucleotide probe was synthesized to be complementary to the coding sequence shown here. The bottom line shows the corresponding coding sequence derived from the 20-kD CaBP cDNA. The nucleotide sequences differed only at five residues (\*).

both orientations. Both strands of pCaBP-4 were completely sequenced with the strategy shown in Fig. 2.

### Sequence of the 20-kD CaBP cDNA and of Its Encoded Protein

The nucleotide sequence together with the deduced amino acid sequence of pCaBP-4 is shown in Fig. 3. The cDNA is 1,052 bases long and contains an open reading frame of 510 bases with an ATG codon 46 bases from the 5' end of the clone and 9 bases downstream from an inframe stop codon. The putative methionine initiation codon is preceded at position -3 (3 nucleotides upstream of the ATG) by an A residue. A purine at this position has been shown to have a dominant effect on selection of a functional initiation codon (16).

The open reading frame that ends with a TAA stop codon would encode a polypeptide of 169 amino acids with a calculated molecular mass of 19,459 D. This value is in good agreement with the estimated molecular mass of the 20-kD CaBP isolated from *Chlamydomonas* (13). The amino acid sequence deduced for the open reading frame is in perfect agreement with internal peptide sequences previously determined for the protein (13). These sequences are underlined in Fig. 3. In addition, the amino acid composition calculated from the sequence data corresponded well to the experimentally determined values for the 20-kD CaBP (Table I). Those differences that were observed may be due to experimental

problems associated with amino acid compositional analysis of gel-eluted material.

Analysis of the predicted amino acid sequence identifies two consensus sequences for potential modification of the protein. The cAMP-dependent protein kinases have been shown to have a strong preference for serines and threonines that are located 2 residues to the carboxy-terminal side of two basic amino acids (17). One such potential cAMP-dependent phosphorylation site is present at serine-167. In addition, the predicted amino acid sequence includes an Asn-Leu-Thr beginning at amino acid residue 133 which matches the consensus sequence derived for the attachment of N-linked carbohydrates to asparagine (14).

It should be noted that the biased use of G or C residues in the third position of all codons, which has been previously reported for several *Chlamydomonas* nuclear-encoded genes (10, 24, 39), was also found for the coding sequence of the 20-kD CaBP cDNA. The frequencies that each of the bases were found in the third position of the codons were (C) 44.1%, (G) 43.5%, (T) 9.4%, and (A) 2.9%. This biased usage of codons is further documented when the sequence of the 59-mer synthetic oligonucleotide probe is compared with the corresponding sequence in the cDNA. As shown in Fig. 1, the sequences differed only in 5 residues, even though the oligonucleotide was unique except for alternative G or C residues in three positions.

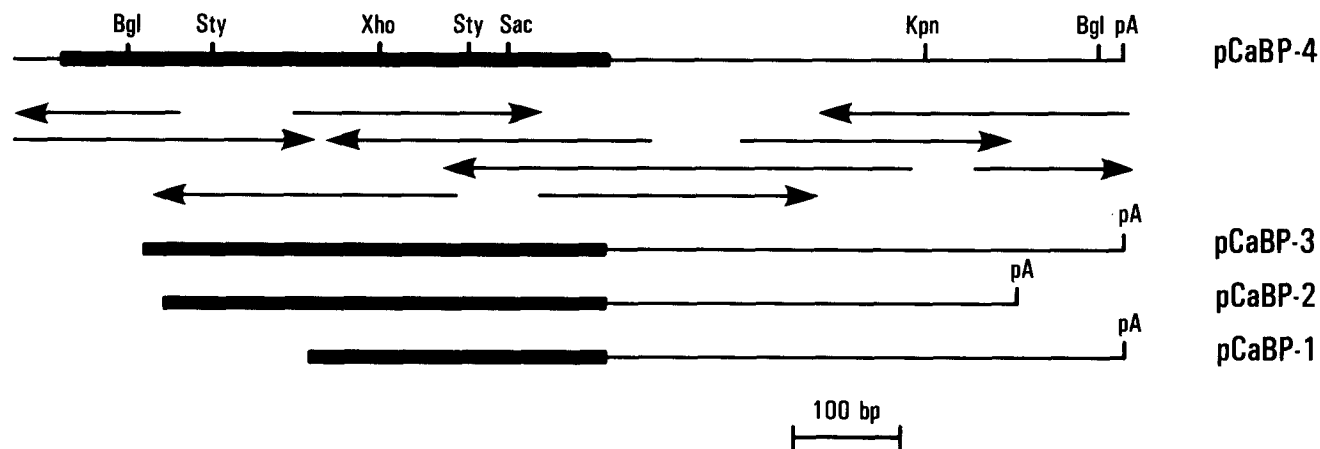


Figure 2. A schematic representation of four overlapping cDNA clones for mRNA encoding the *Chlamydomonas* 20-kD CaBP. The bold lines indicate the protein coding regions, while the thin lines represent the 5' and 3' untranslated regions of the mRNA. A partial restriction endonuclease map and the sequencing strategy used for determining the complete cDNA sequence of pCaBP-4 are shown. *Bgl*, *Bgl* I; *Sty*, *Sty* I; *Xho*, *Xho* I; *Sac*, *Sac* I; *Kpn*, *Kpn* I; *pA*, site of polyadenylation.

50

GCCTGATCATTATCAAAACCGTCTAGCTTCTAAGTCGAAAAC ATG AGC TAC AAG GCA AAG ACC GTC GTC TCC  
Met Ser Tyr Lys Ala Lys Thr Val Val Ser 10

100

GCT CGT CGG GAC CAG AAG AAG GGC CGC GTT GGC CTG ACT GAG GAG CAG AAG CAG GAG ATC CGG GAG  
Ala Arg Arg Asp Gln Lys Lys Gly Arg Val Gly Leu Thr Glu Glu Gln Lys Gln Glu Ile Arg Glu 32

150

GCA TTC GAC CTC TTC GAT ACC GAT GGC ACC GGC ACC ATC GAC GCC AAG GAG CTG AAG GTG GCC ATG  
Ala Phe Asp Leu Phe Asp Thr Asp Gly Ser Gly Thr Ile Asp Ala Lys Glu Leu Lys Val Ala Met 54

200

CGT GCT CTC GGC TTC GAG CCG AAG AAG GAG GAG ATC AAG AAG ATG ATC TCG GAG ATT GAC AAG GAC  
Arg Ala Leu Gly Phe Glu Pro Lys Lys Glu Glu Ile Lys Lys Met Ile Ser Glu Ile Asp Lys Asp 76

250

GGC TCG GGC ACC ATC GAC TTT GAG GAG TTC CTG ACC ATG ATG ACC GCC AAG ATG GGC GAG CGT GAC  
Gly Ser Gly Thr Ile Asp Phe Glu Glu Phe Leu Thr Met Met Thr Ala Lys Met Gly Glu Arg Asp 98

300

TCT CGA GAG GAG ATT CTG AAG GCG TTC CGG CTC TTC GAC GAC GAC AAC TCG GGC ACC ATC ACC ATC  
Ser Arg Glu Glu Ile Leu Lys Ala Phe Arg Leu Phe Asp Asp Asp Asn Ser Gly Thr Ile Thr Ile 120

350

AAG GAC CTC CGG CGT GTG GCC AAG GAG CTG GGC GAG AAC CTG ACT GAG GAG GAG CTC CAA GAG ATG  
Lys Asp Leu Arg Arg Val Ala Lys Glu Leu Gly Glu Asn Leu Thr Glu Glu Leu Gln Glu Met 142

400

ATT GCG GAG GCG GAC CGC AAC GAC GAC AAC GAG ATC GAC GAG GAC GAG TTC ATC CGG ATC ATG AAG  
Ile Ala Glu Ala Asp Arg Asn Asp Asp Asn Glu Ile Asp Glu Asp Glu Phe Ile Arg Ile Met Lys 164

450

500

AAG ACC TCG CTC TTC TAA ACGACGGGGTGGACTGGGCCTATCCATCTTGACGCGGGGCTTGCTTTCACCGCCTGGGGA  
Lys Thr Ser Leu Phe \*\*\*

550

TGTGAGAGTGGTCTCATGCAAATTGAGAGCAATGCGTTGGTCAAGCAGCCGATGATGGGCGCGTGCCGAGCATTCTCAGTATAG  
650

600

ATTATTGACACGGACGCTACCGGTCTGTTCCACCAGTACCACCGCACGCTGTGGGTATGGGTGCTGTGAAGGGTGACACCCCTGTGG  
750

650

TGGCGGACCGATGGCTTTACATTTATGCCCTGATCTTCTAGCTTTTCGTCTTCAGGTACCAAGCAGCGTGGTACAACCTGTACAGAATAC  
800

700

GCATGCCAACCATGCGCGGAGCGAGGTGTTAGACGACACTGTGACTTCGACAGGGAACACACGTTCCGGTTGACAGCAGCACTTG  
850

750

ACTGGATGGACCTGACCAGACTGACATCGAAAGTGCGCTCACAGGCGCTAACATGTAATTTTATCCATAAAAAAAAAAAAAAAAAA  
900

800

850

900

950

1000

1050

**Figure 3.** The nucleotide sequence of the *Chlamydomonas* 20-kD CaBP cDNA. The complete sequence of the pCaBP-4 cDNA is presented together with the deduced amino acid sequence of the coding region. The underlined regions of protein sequence correspond exactly to experimentally determined sequences for the *Chlamydomonas* 20-kD CaBP (13). The numbers on the right side refer to amino acid positions. The arrowhead identifies the site of polyadenylation in pCaBP-2. The two pentanucleotide sequences underlined with wavy lines represent potential polyadenylation signals.

**Table 1. Comparison of the Predicted Amino Acid Composition of the Protein Encoded by pCaBP-4 and the Experimentally Determined Composition of the *Chlamydomonas* Basal Body-associated Calcium-Binding Protein**

Amino acid	pCaBP-4 predicted*	<i>Chlamydomonas</i> 20-kD CaBP <sup>†</sup>
Ala	11	14.4
Arg	12	12.9
Asp	22	19.9
Cys	0	ND
Glu	28	25.8
Gly	10	14.3
His	0	1.0
Ile	13	11.6
Leu	12	14.5
Lys	18	14.1
Met	8	5.8
Phe	9	8.9
Pro	1	4.2
Ser	8	8.6
Thr	11	10.8
Trp	0	ND
Tyr	1	2.6
Val	5	7.7

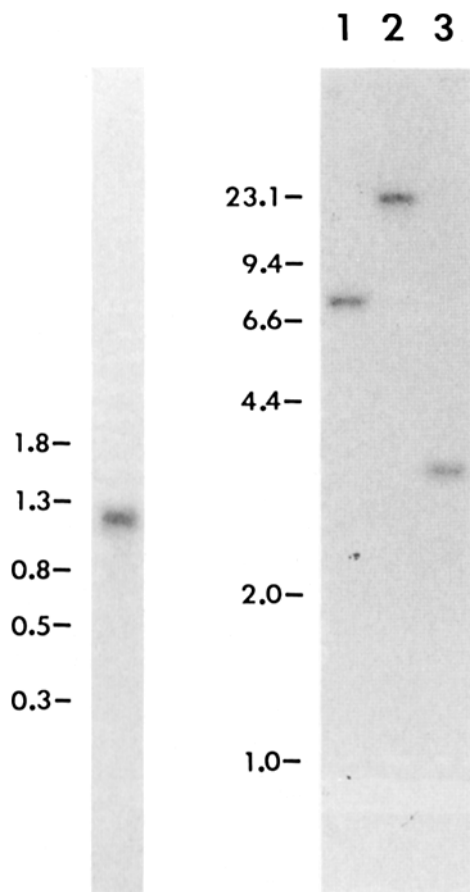
\* Values are residues per molecule calculated from the predicted amino acid sequence.

† Values are mole per 20,000 g as previously estimated from analysis of gel-eluted protein (13).

The 20-kD CaBP cDNA clone pCaBP-4 contains an extended 3' untranslated region of 496 bases that includes 16 polyadenylic acid residues. 14 bases upstream from the beginning of the poly(A) tract is the pentanucleotide TGTA (Fig. 3, wavy lines), which has previously been identified as a potential polyadenylation signal in *C. reinhardtii* (30). As shown in Fig. 2, pCaBP-1 and pCaBP-3 also contained 3' untranslated regions of the same length as pCaBP-4 and were found by sequence analysis to be polyadenylated at or near the same residue. In contrast, pCaBP-2 contained a truncated 3' untranslated region. DNA sequencing revealed that pCaBP-2 was polyadenylated at the nucleotide corresponding to nucleotide 938 in pCaBP-4 (Fig. 3, arrowhead). 14 nucleotides upstream from this site is the pentanucleotide TGTGA (Fig. 3, wavy lines) which may represent the polyadenylation signal used to generate the RNA species represented by pCaBP-2.

#### Characterization of the 20-kD CaBP RNA and Analysis of Genomic DNA

Poly(A) RNA was isolated from *Chlamydomonas* and subjected to blot hybridization using <sup>32</sup>P-labeled pCaBP-2 cDNA as a probe (Fig. 4 left). A single size transcript of ~1.1 kb was detected. The same RNA blot was also probed with radiolabeled pCaBP-4 cDNA and the results were identical to those shown in Fig. 4 left (data not shown). Total genomic DNA was isolated from *C. reinhardtii* and after



**Figure 4.** RNA and genomic DNA blot analysis. (Left) Poly(A) RNA (2  $\mu$ g) from *Chlamydomonas* cells was subjected to electrophoresis on a denaturing formaldehyde agarose gel, blotted, and then hybridized to  $^{32}$ P-labeled pCaBP-2. (Right) *Chlamydomonas* genomic DNA was digested with each of three restriction endonucleases, subjected to electrophoresis, blotted, and hybridized to  $^{32}$ P-labeled pCaBP-2. Lane 1, Bam HI; Lane 2, Eco RI; Lane 3, Pst I.  $^{32}$ P-labeled RNAs and DNAs of known sizes were used as molecular weight markers, and their migration positions are marked in kilobases at the left of each gel.

digestion with different restriction enzymes, the fragments were examined by Southern blot analysis with cDNA from pCaBP-2. As shown in Fig. 4, right, lanes 1–3, a single hybridizing fragment was detected in each of the three digests. The same hybridizing fragments were detected when the blot was probed with cDNA from pCaBP-4 (data not shown).

#### **Protein Sequence Relatedness of the 20-kD CaBP with Calmodulin and the Yeast *CDC31* Gene Product**

In previous studies on the 20-kD CaBP isolated from *Chlamydomonas*, evidence was obtained based on its electrophoretic behavior, amino acid composition, peptide sequencing, and antigenic properties that the protein was related to calmodulin but clearly distinct from it (13). With the deduced complete amino acid sequence for the protein, the extent of sequence homology to calmodulin and other calcium-binding proteins could be evaluated. Not surprisingly, a computer search revealed a significant sequence-relatedness of the 20-kD CaBP with calmodulin from *Chlamydo-*

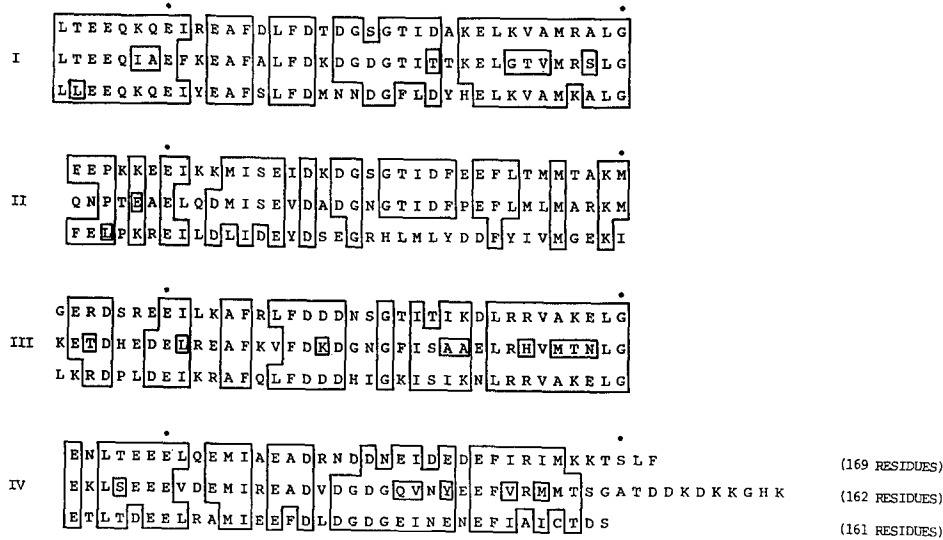
*monas* and other organisms (45–48%) and rabbit skeletal muscle troponin C (41%). Lower degrees of similarity were found to other well-characterized calcium-binding proteins such as rabbit skeletal muscle myosin catalytic light chains (27%). Strikingly, the 20-kD CaBP showed the highest sequence identity (50%) with the deduced amino acid sequence of the yeast, *Saccharomyces cerevisiae CDC31* gene product. This gene, which was cloned by complementation, was reported to contain a significant homology with calmodulin in its coding sequence (3).

Fig. 5 contains a sequence comparison of the *Chlamydomonas* 20-kD CaBP with the experimentally determined sequence for *Chlamydomonas* calmodulin (22) and the deduced amino acid sequence of the yeast *CDC31*-encoded protein (3). When leucine-22 of the *Chlamydomonas* 20-kD CaBP is aligned with leucine-7 of *Chlamydomonas* calmodulin and leucine-17 of the yeast *CDC31* gene product, a linear homology with no gaps or insertions is found to extend to the carboxy termini of the proteins. In Fig. 5 only identical residues are boxed. It should be noted that *Chlamydomonas* calmodulin is longer than all other calmodulins characterized to date due to the presence of an additional 3 amino acids at the amino terminus and 11 amino acids at the carboxy terminus of the protein. Except for this feature, *Chlamydomonas* calmodulin differs from vertebrate calmodulins in only 15 residues, most of which are conservative differences (22).

The 20-kD CaBP and the yeast *CDC31* gene product both contain amino-terminal domains which are not found in calmodulin from *Chlamydomonas* or other organisms. In the 20-kD CaBP, the amino-terminal domain consists of 21 amino acids. This domain contains a concentration of positively charged amino acids and the only tyrosine in the protein is located at the third residue from the DNA-encoded amino terminus. The amino-terminal domain of the yeast *CDC31* gene product is shorter and bears no apparent sequence relatedness to the amino-terminal domain of the 20-kD CaBP. While the *Chlamydomonas* 20-kD CaBP showed a 45 and 50% sequence identity, respectively, with *Chlamydomonas* calmodulin and the yeast *CDC31* gene product, the protein sequence identity between the latter two proteins was less extensive (36%).

The linear sequence homology of the 20-kD CaBP with calmodulin extends through the four recognized calcium-binding domains of calmodulin (Fig. 5, I–IV). In Fig. 5, the sequences corresponding to these domains lie on each line between the residues indicated with a dot above them. An analysis of these sequences in the 20-kD CaBP (Fig. 6) reveals the pattern of conserved residues anticipated for calcium-binding domains of EF hand structure. An EF hand is a linear sequence of  $\sim 30$  amino acids where two nearly perpendicular  $\alpha$ -helices flank a 12-residue calcium-binding loop (for a review see reference 18). In Fig. 6 residues 1–9 constitute the E  $\alpha$ -helical region, residues 10–21 make up the calcium-binding loop, and residues 22–30 constitute the F  $\alpha$ -helical region. In the 20-kD CaBP, as in calmodulin and troponin C, a glutamic acid is found at position 1 which begins the first helix of the EF hand. The flanking  $\alpha$ -helices in calcium-modulated proteins have been proposed to be amphipathic (18). In the 20-kD CaBP those residues predicted to be on the hydrophobic face of the helices (labeled *n*) are almost uniformly nonpolar. In the potential calcium-binding

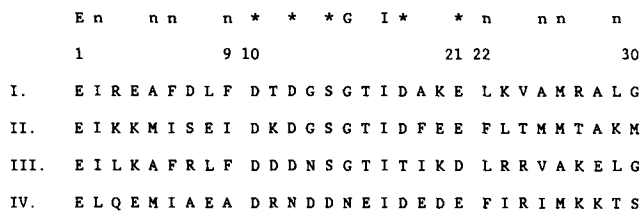
CHLAMYDOMONAS 20 kD CaBP - M S Y K A K T V V S A R R D Q K K G R V G  
 CHLAMYDOMONAS CALMODULIN - A A N T E Q  
 YEAST CDC31 GENE PRODUCT - M S K N R S S L Q S G P L N S E



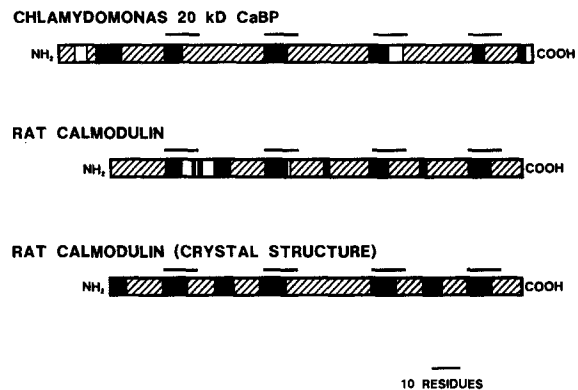
**Figure 5.** Comparison of the amino acid sequence of the *Chlamydomonas* 20-kD CaBP with *Chlamydomonas* calmodulin (22) and the yeast *Saccharomyces cerevisiae* CDC31-encoded protein (3). Only identical residues are boxed. Roman numerals to the left denote the four calcium-binding domains in calmodulin; the amino acids that constitute the EF hand calcium-binding domains lie between the residues identified with a dot above them.

loops, the residues found at the positions predicted to bind calcium in calcium-modulated proteins are all amino acids with oxygen-containing side chains in the 20-kD CaBP. The glycine at position 15, which is believed to occur at a sharp bend in the calcium-binding loop in calcium-modulated proteins, is conserved in the 20-kD CaBP except in domain IV where an asparagine residue is found. The isoleucine at position 17 which is proposed to contribute a hydrophobic side chain to the core of the calcium-modulated proteins is also conserved in the 20-kD CaBP.

The Robson algorithm (8) was used to predict the second-



**Figure 6.** The *Chlamydomonas* 20-kD CaBP contains four potential calcium-binding domains of predicted EF hand structure. The amino acid sequences in the 20-kD CaBP corresponding to the four EF hand domains (I-IV) of calmodulin are shown. The EF hand model position numbers are from Kretsinger (18) with residues 1-9, 10-21, and 22-30 corresponding respectively to the E helix, calcium-binding loop, and F helix of the calcium-binding domains. The amino acids found at positions 2, 5, 6, 9, and 22, 25, 26, 29 (denoted n), which are predicted to be on the hydrophobic face of the  $\alpha$ -helices in calcium-modulated proteins, are almost uniformly nonpolar in the 20-kD CaBP. The residues found at the positions predicted to ligand calcium in the calcium-modulated proteins (denoted \*) are all amino acids with oxygen-containing side chains in the 20-kD CaBP. The glutamic acid (E) at position 1, the glycine (G) at position 15, and the isoleucine (I) at position 17 in the calcium-modulated proteins are also conserved in the 20-kD CaBP except at position 15 in domain IV where an asparagine (N) instead of glycine (G) is found.



**Figure 7.** The predicted secondary structure of the *Chlamydomonas* 20-kD CaBP conforms to the helix-loop-helix (or EF hand) arrangement for the potential calcium-binding domains. The Robson algorithm (8) was used to predict the secondary structure of the *Chlamydomonas* 20-kD CaBP and vertebrate calmodulin. The secondary structure of vertebrate calmodulin as deduced from x-ray crystallography (2) is also shown. Hatched boxes,  $\alpha$ -helical segments; unfilled boxes,  $\beta$ -sheet; filled boxes,  $\beta$ -turn or random coil conformations. Potential calcium-binding loops are indicated (—).

ary structure of the 20-kD CaBP. As shown in Fig. 7, a helix-loop-helix (EF hand) structure for the potential calcium-binding regions was predicted for the protein. Although the actual secondary structure of calmodulin, as deduced from the x-ray crystal structure (2), differed in details from the structure predicted by the Robson algorithm (Fig. 7), the analysis confirms that the *Chlamydomonas* 20-kD CaBP has a structural organization similar to that found in calmodulin and other sequence-related calcium-modulated proteins.

### Discussion

The results presented in this study indicate that a cDNA con-

taining the entire coding region for a *Chlamydomonas* basal body-associated 20-kD CaBP has been identified and characterized. The predicted size and amino acid composition of the protein encoded by pCaBP-4 showed a close correspondence with those experimentally determined for the 20-kD CaBP (13). In addition, a perfect match was found in the deduced protein sequence of pCaBP-4 with partial amino acid sequences derived from the isolated 20-kD CaBP (13). These data provide strong evidence that pCaBP-4 contains the entire coding region for the 20-kD CaBP and that the deduced amino acid sequence is the actual sequence of the protein.

The results obtained by Northern and Southern blot analyses indicate that the 20-kD CaBP is encoded by a single-copy gene from which is derived a single size class transcript. pCaBP-4 contains most, if not all, of the sequence for the 20-kD CaBP RNA since there is a close similarity in the size of the 20-kD CaBP RNA, as determined by Northern blot analysis, and the length of the cDNA, as determined by sequence analysis. Further proof of the full-length nature of the clone will require the isolation and characterization of genomic clones. In addition to pCaBP-4, three polyadenylated partial cDNA clones for the 20-kD CaBP were isolated. Two of these clones contained 3' untranslated regions of the same length (480 bp) as that found in pCaBP-4 and were polyadenylated at or near the same residue. One clone was identified with a truncated 3' untranslated region of 381 bp in length. These data indicate that the 20-kD CaBP RNA is subject to alternative polyadenylation. Although alternative polyadenylation of mRNAs in *Chlamydomonas* has not been previously reported, it has been observed recently for another *Chlamydomonas* gene, the nuclear-encoded photosystem II OEE 1 protein (Mayfield, S., Research Institute of Scripps Clinic, personal communication).

The deduced amino acid sequence of the 20-kD CaBP was found to contain a significant linear sequence relatedness to the well-characterized calcium-modulated proteins, calmodulin and troponin C. Analysis of its primary protein sequence and predicted secondary structure indicate that the 20-kD CaBP is likely to contain four homologous calcium-binding domains which conform to the helix-loop-helix or EF hand structure found in calcium-modulated proteins (18). The 20-kD CaBP appears to be more closely related to calmodulin than troponin C; the position of the four potential calcium-binding domains are in exact sequence register to that found in calmodulin (see Fig. 7). Similar to calmodulin (36), the first and second domains are more closely related in sequence, respectively, to domains 3 and 4 than are other pairs of domains (see Fig. 6). In addition, the 20-kD CaBP contains no cysteine or tryptophan residues, a feature of most calmodulins characterized to date (for a recent review, see reference 25). Calmodulin has been shown to interact in a calcium-dependent manner with a large number of receptor proteins and drugs (for reviews see references 15, 34). Given the sequence relatedness of the 20-kD CaBP with calmodulin, the protein may share, in addition to calcium-binding, some of the biochemical activities associated with calmodulin.

Until recently, the characterization of the calcium-binding properties and the *in vitro* biochemical and biophysical properties of the 20-kD CaBP was prohibited by the microgram quantities of purified protein available for analysis (13). However, the cloning of cDNA for the 20-kD CaBP, which

has been reported here, has resolved this problem. The cDNA insert in pCaBP-4 has been subcloned into a bacterial expression vector and the 20-kD CaBP, as an unfused product, has been expressed and purified from bacterial cultures in milligram quantities (Huang, B., and A. Mengersen, unpublished results). Detailed biochemical and structural studies on the protein are now possible, and are being pursued.

The *Chlamydomonas* 20-kD CaBP has been shown to be a component of calcium-sensitive contractile fibers that link the basal bodies of the complex to each other and the complex as a whole to the nucleus (13, 38). It has been observed with antibodies raised against the homologous protein from the alga *Tetraselmis* that antigenic determinants related to the algal 20-kD CaBPs are expressed in association with the centrosomes of mammalian cells in culture (28). In this study, a significant protein sequence identity was found between the *Chlamydomonas* 20-kD CaBP and the yeast *S. cerevisiae* cell division cycle-related gene product *CDC31* (3). Although the *CDC31*-encoded protein has not yet been isolated or its cellular localization determined, temperature-sensitive alleles for this gene show a cell division cycle arrest related to a defect in the duplication of the spindle pole body (4). The spindle pole body in yeast cells is a structure found associated with the nuclear envelope and it functions as the major microtubule-organizing center of the cell. As such, it is the yeast homologue to the basal body complex in *Chlamydomonas* and the centrosome in animal cells. The association of these sequence-related proteins to microtubule-organizing centers of divergent structure suggests that the proteins may be functionally related. To investigate this possibility, complementation analysis of *CDC31* mutants transformed with the *Chlamydomonas* 20-kD CaBP cDNA incorporated into a stable, inducible yeast expression vector is currently being pursued.

Previous studies have suggested that the 20-kD CaBP may be a structural component of the algal basal body-associated contractile fibers (27). Analysis of the primary sequence of the 20-kD CaBP revealed the presence of an amino-terminal domain of 21 amino acids that is not found in any of the well-characterized calcium-modulated proteins. The yeast *CDC31*-encoded gene product also contains an amino-terminal extension, but with no apparent relatedness to the amino-terminal domain of the 20-kD CaBP. Although there is no obvious mechanism, based on the primary sequence of this domain in the 20-kD CaBP, which might contribute to the potential cytoskeletal properties of the protein, the domain does represent the major divergent feature of the protein from other calcium-modulated proteins that are known to have primarily regulatory rather than structural functions. Future studies on the structure of the 20-kD CaBP may reveal a role for the amino-terminal domain which is not obvious at this time.

There is evidence to suggest that the 20-kD CaBP in *Chlamydomonas* may be subject to co- or posttranslational modification. Two isoforms of the protein, which differ in their pIs, have been identified in the cell (38). Although the relationship of these two isoforms has not been determined, it has been demonstrated that the homologous protein in the alga *Tetraselmis* is phosphorylated *in vivo* (27). A consensus sequence for cAMP-dependent protein kinase phosphorylation of serine-167 in the deduced amino acid sequence of the 20-kD CaBP was identified. Whether this site is used *in vivo*



remains to be determined. A consensus sequence for the attachment of N-linked carbohydrates was also identified in the protein sequence of the 20-kD CaBP. However, since there is no evidence that the 20-kD CaBP is a membrane component or a secreted product that would transit the Golgi complex, it is unclear that this site is of physiological significance. It is curious, however, that the identified site of potential glycosylation in the 20-kD CaBP is asparagine-133. This amino acid position corresponds in the linear sequence homology of the 20-kD CaBP with calmodulin to vertebrate calmodulin lysine-115, which has been shown to contain the unusual modification of trimethylation (36).

All of the available data suggest that the basal body-associated 20-kD CaBP is ancestrally related to members of the calcium-modulated family of proteins. Genomic sequences encoding the *Chlamydomonas* 20-kD CaBP have been cloned and introns have been found in the gene (Lee, V. D., and B. Huang, unpublished results). The number of introns and the size and position of the exons are being determined by DNA sequencing of the genomic clones. A similar analysis of the genomic organization of the *Chlamydomonas* calmodulin gene is also being pursued (Zimmer, W. E., and D. M. Watterson, personal communication). A comparison of the results obtained from these two studies is likely to reveal important insights into the evolutionary relationship of the 20-kD CaBP with calmodulin and other members of the calcium-modulated family of proteins.

Since the 20-kD CaBP, characterized in this and the accompanying paper (13), is clearly a novel calcium-binding protein associated with a calcium-sensitive contractile fiber system, the term caltractin is proposed for the protein.

We thank Nalin Kumar and Matt Haynes for their generous advice and assistance during all phases of this work; D. Martin Watterson for his critical reading of the manuscript; and Cheryl Negus for typing the manuscript.

This research was supported by a grant to B. Huang from the National Institutes of Health (GM-38113).

Received for publication 28 January 1988, and in revised form 11 March 1988.

## References

- Ares, Jr., M., and Howell, S. H. 1982. Cell cycle stage-specific accumulation of mRNAs encoding tubulin and other polypeptides in *Chlamydomonas*. *Proc. Natl. Acad. Sci. USA*. 79:5577-5581.
- Babu, Y. S., J. S. Sack, T. J. Greenhough, C. E. Bugg, A. R. Means, and W. J. Cook. 1985. Three dimensional structure of calmodulin. *Nature (Lond.)*. 315:37-40.
- Baum, P., C. Furlong, and B. Byers. 1986. Yeast gene required for spindle pole body duplication: homology of its gene product with Ca<sup>2+</sup>-binding proteins. *Proc. Natl. Acad. Sci. USA*. 83:5512-5516.
- Byers, B. 1981. Multiple roles of the spindle pole bodies in the life cycle of *Saccharomyces*. In *Molecular Genetics in Yeast*. Vol. 17. D. Von Wettstein, J. Friis, M. Keilland-Brandt, and A. Stenderup, editors. Munksgaard, Copenhagen, 119-131.
- Davis, L. G., M. D. Dibner, and J. F. Battey. 1986. *Methods in Molecular Biology*. Elsevier Science Publishing Co., Inc., New York. 144 pp.
- Dedman, J. R., R. L. Jackson, W. E. Schreiber, and A. R. Means. 1978. Sequence homology of the Ca<sup>2+</sup>-dependent regulator of cyclic nucleotide phosphodiesterase from rat testis with other Ca<sup>2+</sup>-binding proteins. *J. Biol. Chem.* 253:343-346.
- Feinberg, A., and B. Vogelstein. 1984. Addendum: a technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 137:266-267.
- Garnier, J., D. J. Osguthorpe, and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* 120:97-120.
- Glisin, V., R. Crkvenjakov, and C. Byus. 1974. Ribonucleic acid isolated by cesium chloride centrifugation. *Biochemistry*. 13:2633-2637.
- Goldschmidt-Clermont, M., and M. Rahire. 1986. Sequence, evolution and differential expression of the two genes encoding variant small subunits of ribulose biphosphate carboxylase/oxygenase in *Chlamydomonas reinhardtii*. *J. Mol. Biol.* 191:421-432.
- Helms, C., M. Y. Graham, J. E. Dutchik, and M. V. Olson. 1985. A new method for purifying lambda DNA from phage lysates. *DNA*. 4:39-49.
- Herzberg, O., and M. N. G. James. 1985. Structure of the calcium regulatory muscle protein troponin-C at 2.8 Å resolution. *Nature (Lond.)*. 313:653-659.
- Huang, B., D. M. Watterson, V. D. Lee, and M. J. Schibler. 1988. Purification and characterization of a basal body-associated Ca<sup>2+</sup>-binding protein. *J. Cell Biol.* 107:121-131.
- Hubbard, S. C., and R. J. Ivatt. 1981. Synthesis and processing of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* 50:555-583.
- Klee, C. B., and T. C. Vanaman. 1982. Calmodulin. *Adv. Protein Chem.* 35:213-321.
- Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*. 44:283-292.
- Krebs, E. E., and J. A. A. Beavo. 1979. Phosphorylation-dephosphorylation of enzymes. *Annu. Rev. Biochem.* 48:923-959.
- Kretsinger, R. H. 1980. Structure and evolution of calcium-modulated proteins. *CRC Crit. Rev. Biochem.* 8:119-174.
- Kretsinger, R. H., and C. C. Barry. 1975. The predicted structure of the calcium-binding component of troponin. *Biochim. Biophys. Acta*. 405:40-52.
- Kretsinger, R. H., S. E. Rudnick, and L. J. Weissman. 1986. Crystal structure of calmodulin. *J. Inorg. Biochem.* 28:289-302.
- Lefebvre, P. A., C. D. Silflow, E. D. Wieben, and J. L. Rosenbaum. 1980. Increased levels of mRNAs for tubulin and other flagellar proteins after amputation or shortening of *Chlamydomonas* flagella. *Cell*. 20:469-477.
- Lukas, T. J., M. E. Wiggins, and D. M. Watterson. 1985. Amino acid sequence of a novel calmodulin from the unicellular alga *Chlamydomonas*. *Plant. Physiol. (Bethesda)*. 78:477-483.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Mayfield, S. P., M. Rahire, G. Frank, H. Zuber, and J.-D. Rochaix. 1987. Expression of the nuclear gene encoding oxygen-evolving enhancer protein 2 is required for high levels of photosynthetic oxygen evolution in *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. USA*. 84:749-753.
- Roberts, D. M., T. J. Lukas, and D. M. Watterson. 1986. Structure, function, and mechanism of action of calmodulin. *CRC Crit. Rev. Plant Sci.* 4:311-339.
- Rochaix, J. D. 1980. Restriction fragments from *Chlamydomonas* chloroplast DNA. *Methods Enzymol.* 65:785-795.
- Salisbury, J. L., A. T. Baron, D. E. Coling, V. E. Martindale, and M. A. Sanders. 1986. Calcium-modulated contractile proteins associated with the eucaryotic centrosome. *Cell Motil. Cytoskeleton*. 6:193-197.
- Salisbury, J. L., A. T. Baron, B. Surek, and M. Melkonian. 1984. Striated fiber roots: isolation and partial characterization of a calcium-modulated contractile proteins. *J. Cell Biol.* 99:962-970.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*. 74:5463-5467.
- Silflow, C. D., R. L. Chisholm, T. W. Conner, and L. P. W. Ranum. 1985. The two alpha-tubulin genes of *Chlamydomonas reinhardtii* code for slightly different proteins. *Mol. Cell. Biol.* 5:2389-2398.
- Sundaralingam, M., R. Bergstrom, G. Strasburg, S. T. Rao, P. Roychowdhury, M. Greaser, and B. C. Wang. 1985. Molecular structure of troponin C from chicken skeletal muscle at 3-angstrom resolution. *Science (Wash. DC)*. 227:945-948.
- Szebenyi, D. M. E., S. K. Obendorf, and K. Moffat. 1981. Structure of vitamin D-dependent calcium-binding protein from bovine intestine. *Nature (Lond.)*. 294:327-332.
- Tabor, S., and C. C. Richardson. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA*. 84:4767-4771.
- Van Eldik, L. J., and D. M. Watterson. 1985. Calmodulin structure and function. In *Calcium and Cell Physiology*. D. Marme, editor. Springer-Verlag New York Inc., Berlin, Federal Republic of Germany. 105-126.
- Van Eldik, L. J., J. G. Zengdegi, F. R. Marshak, and D. M. Watterson. 1982. Calcium-binding proteins and the molecular basis of calcium action. *Int. Rev. Cytol.* 77:1-61.
- Watterson, D. M., F. Sharief, and R. C. Vanaman. 1980. The complete amino acid sequence of the Ca<sup>2+</sup>-dependent modulator protein of bovine brain. *J. Biol. Chem.* 255:962-975.
- Whitehead, A. S., G. Goldberger, D. Woods, A. F. Markham, and H. R. Colten. 1983. Use of a cDNA clone for the fourth component of human complement (C4) for analysis of a genetic deficiency of C4 in guinea pig. *Proc. Natl. Acad. Sci. USA*. 80:5387-5391.
- Wright, R. L., J. L. Salisbury, and J. W. Jarvik. 1985. The nucleus/basal body connector in *Chlamydomonas reinhardtii* that may function in basal body localization or segregation. *J. Cell Biol.* 101:1903-1912.
- Youngblom, J., J. A. Schloss, and C. D. Silflow. 1984. The two β-tubulin genes of *Chlamydomonas reinhardtii* code for identical proteins. *Mol. Cell. Biol.* 4:2688-2696.