Gardell, S. J., Craik, C. S., Hilvert, D., Urdea, M. S., & Rutter, W. J. (1985) *Nature 317*, 551–555.
Gardell, S. J., Hilvert, D., Barnett, J., Kaiser, E. T., & Rutter, W. J. (1987) *J. Biol. Chem. 262*, 576–582.
Hand, E. S., & Jencks, W. P. (1975) *J. Am. Chem. Soc. 97*, 6221–6230.
Jarabak, R., Colvin, M., Moolgavkar, S. H., & Talalay, P. (1969) *Methods Enzymol. 15*, 642–666.
Jones, J. B., & Wigfield, D. C. (1969) *Can. J. Chem. 47*, 4459–4466.
Kawahara, F. S., & Talalay, P. (1960) *J. Biol. Chem. 235*, PC1–PC2.
King, E. L., & Altman, C. (1956) *J. Phys. Chem. 60*, 1375–1378.
Knowles, J. R. (1987) *Science 236*, 1252–1258.
Kuliopulos, A., Shortle, D., & Talalay, P. (1987a) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8893–8897.
Kuliopulos, A., Westbrook, E. M., Talalay, P., & Mildvan, A. S. (1987b) *Biochemistry 26*, 3927–3937.
Kuliopulos, A., Mildvan, A. S., Shortle, D., & Talalay, P. (1989) *Biochemistry 28*, 149–159.
Kuliopulos, A., Talalay, P., & Mildvan, A. S. (1990) *Biophys. J. 57*, 39a.
Malhotra, S. K., & Ringold, H. J. (1965) *J. Am. Chem. Soc. 87*, 3228–3236.
Mildvan, A. S., & Weiner, H. (1969) *Biochemistry 8*, 552–562.
Mildvan, A. S., & Engle, J. L. (1972) *Methods Enzymol. 26*, 654–682.
Mullen, G. P., Shenbagamurthi, P., & Mildvan, A. S. (1989) *J. Biol. Chem. 264*, 19637–19647.
O'Leary, M. (1989) *Annu. Rev. Biochem. 58*, 377–401.
Perera, S. K., Dunn, W. A., & Fedor, L. R. (1980) *J. Org. Chem. 45*, 2816–2821.
Rebek, J., Jr. (1988) *J. Mol. Recognit. 1*, 1–8.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Serpersu, E. H., Shortle, D., & Mildvan, A. S. (1987) *Biochemistry 26*, 1289–1300.
Weber, D. J., Serpersu, E. H., Shortle, D., & Mildvan, A. S. (1990) *Biochemistry 29*, 8632–8642.
Westbrook, E. M. (1976) *J. Mol. Biol. 103*, 659–664.
Xue, L., Talalay, P., & Mildvan, A. S. (1990) *Biochemistry 29*, 7491–7500.
Zoller, M. J., & Smith, M. (1983) *Methods Enzymol. 100*, 468–500.

Amino Acid Sequence of the Cyclic GMP Stimulated Cyclic Nucleotide Phosphodiesterase from Bovine Heart†

Hai Le Trong,‡ Norbert Beier,§,‖ William K. Sonnenburg,§ Steven D. Stroop,§ Kenneth A. Walsh,‡ Joseph A. Beavo,§ and Harry Charbonneau*,‡

Departments of Biochemistry and Pharmacology, University of Washington, Seattle, Washington 98195

Received June 12, 1990; Revised Manuscript Received August 14, 1990

ABSTRACT: The complete amino acid sequence of the cyclic GMP stimulated cyclic nucleotide phosphodiesterase (cGS-PDE) of bovine heart has been determined by analysis of five digests of the protein; placement of the C-terminal 330 residues has been confirmed by interpretation of the corresponding partial cDNA clone. The holoenzyme is a homodimer of two identical N^{α}-acetylated polypeptide chains of 921 residues, each with a calculated molecular weight of 103 244. The C-terminal region, residues 613–871, of the cGS-PDE comprises a catalytic domain that is conserved in all phosphodiesterase sequences except those of PDE I from *Saccharomyces cerevisiae* and a secreted PDE from *Dictyostelium*. A second conserved region, residues 209–567, is homologous to corresponding regions of the α and α' subunits of the photoreceptor phosphodiesterases. This conserved domain specifically binds cGMP and is involved in the allosteric regulation of the cGS-PDE. This regulatory domain contains two tandem, internal repeats, suggesting that it evolved from an ancestral gene duplication. Common cyclic nucleotide binding properties and a distant structural relationship provide evidence that the catalytic and regulatory domains within the cGS- and photoreceptor PDEs are also related by an ancient internal gene duplication.

Cyclic nucleotides mediate a variety of cellular responses to biological stimuli. The concentrations of the cyclic nucleotides are controlled by the opposing action of adenylate or guanylate cyclases and phosphodiesterases (PDEs),[1] which are in turn regulated by transmembrane signals or second-messenger ligands (e.g., calcium/calmodulin or cGMP). Among this milieu of interacting regulatory systems, a variety of cyclic nucleotide phosphodiesterase isozymes (Beavo, 1988; Beavo & Reifsynder, 1990) have been identified; they share the potential to return cyclic nucleotide stimulated systems to the ground state in response to regulatory signals. Different tissues express PDEs with different responsiveness to natural regulatory metabolites. The present study defines the amino acid sequence of one of the larger PDEs, the cGMP-stimulated

† This work was supported by National Institutes of Health Grants GM15731 (to K.A.W.) and DK21723 and EY08197 (to J.A.B.) and in part by Fellowship Be 1003 (to N.B.) from the Deutsche Forschungsgemeinschaft.

* Address correspondence to this author.
‡ Department of Biochemistry.
§ Department of Pharmacology.
‖ Present address: E. Merck, Postfach 4119, 6100 Darmstadt 1, Germany.

[1] Abbreviations: cGS-PDE, cGMP-stimulated cyclic nucleotide phosphodiesterase; HPLC, high-performance liquid chromatography; TFA, trifluoroacetic acid.

0006-2960/90/0429-10280$02.50/0

PDE (cGS-PDE) from bovine heart.

A distinctive feature of this PDE isozyme is the ability of cGMP to stimulate cAMP hydrolysis and its ability to hydrolyze either cGMP or cAMP with positively cooperative kinetics (Beavo et al., 1971; Martins et al., 1982; Wada et al., 1987). Evidence from several different studies shows that the enzyme has a noncatalytic, cGMP-specific site that may account for the stimulation of cAMP hydrolysis by cGMP (Erneux et al., 1981; Martins et al., 1982; Miot et al., 1985; Stroop et al., 1989). At physiological cyclic nucleotide concentrations, this enzyme responds to elevated cGMP concentrations with an enhanced hydrolysis of cAMP. Thus, cGS-PDE allows increases in cGMP concentration to moderate or inhibit cAMP-mediated responses. The primary sequence presented here provides the molecular framework for understanding the regulatory properties and domain substructure of this enzyme and for comparing it with other PDE isozymes that respond to different signals.

Materials and Methods

Cyanogen bromide and citraconic anhydride are products of Eastman Kodak. Hydroxylamine hydrochloride and iodoacetic acid were obtained from Sigma, 4-vinylpyridine was from Aldrich; TPCK-trypsin, chymotrypsin, and *Staphylococcal aureus* V8 protease were from Cooper Biomedical, Worthington, and Miles, respectively. *Pseudomonas fragi* endoproteinase Asp-N was purchased from Boehringer Mannheim; *Achromobacter protease* I was a gift of Dr. T. Masaki (Ibaraki University, Japan).

Preparation of Reduced and Alkylated cGS-PDE. cGS-PDE was purified from bovine heart by using procedures described by Martins et al. (1982) with some modifications by Stroop et al. (1989). The purity of the enzyme was assessed by silver staining after SDS–PAGE.

Purified cGS-PDE was reduced and carboxymethylated as described by Le Trong et al. (1989). S-Pyridylethylated protein was prepared by first dialyzing against 0.25 M Tris-HCl/1 mM EDTA, pH 8.2, and concentrating to 0.5–2.0 mL in a Centricon 10 ultrafiltration tube (Amicon). Guanidine hydrochloride (6.0 M) and 2-mercaptoethanol (about 100-fold molar excess over protein thiols) were then added, and the mixture was incubated at room temperature in the dark under argon for 3 h. Pure 4-vinylpyridine was added in 5-fold molar excess over the thiols, and the mixture was incubated at room temperature under argon for 2 h. The protein solution was dialyzed against 9% formic acid and lyophilized.

Chemical Cleavages. Methionine residues in 17.5 nmol of reduced and carboxymethylated cGS-PDE were cleaved in 70% formic acid by treatment with a hundredfold molar excess of cyanogen bromide for 24 h in the dark, under argon, at room temperature (Gross, 1967). After dilution 15-fold with water, the mixture was lyophilized and fractionated by size on two tandem TSK-3000 SW columns (Figure 1).

An asparaginyl–glycyl bond in 310 pmol of peptide R12 was cleaved in a freshly prepared solution of 2 M hydroxylamine, in 6 M guanidine hydrochloride and 0.2 M potassium carbonate, pH 9.0 (Bornstein & Balian, 1977). The reaction mixture was incubated at 45 °C for 4 h and then separated by reverse-phase HPLC on a C8 column to obtain the pure peptide R12-NG1.

Cleavage of Peptides on a Glass Fiber Filter. Internal sequences of peptides blocked at their amino termini were obtained by chemical cleavage at serine and threonine residues on a glass fiber filter. Peptide that had been analyzed in a gas-phase sequencer was dried on the filter with argon and then exposed to trifluoroacetic acid (TFA) vapors for 10 min

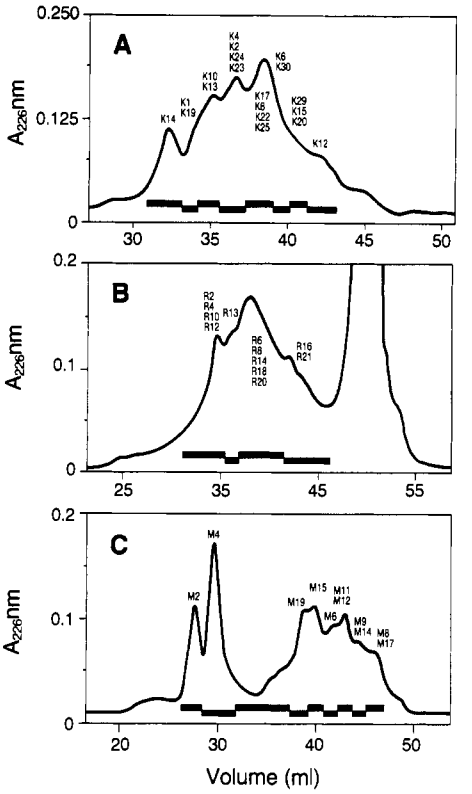

FIGURE 1: Primary separation of mixtures of peptides generated by size-exclusion chromatography after cleavage at lysine (A), arginine (B), and methionine (C) of reduced, alkylated cGS-PDE. The separations were performed in 6 M guanidine hydrochloride/10 mM sodium phosphate, pH 6.0, at a flow rate of 0.5 mL/min on two TSK-3000 SW (7.5 × 600 mm) columns (C) or two TSK-2000 SW (7.5 × 600 mm) columns (A, B) connected in series. Fractions were pooled as indicated and further separated by reverse-phase HPLC. Peptides are listed from the top in order of increasing hydrophobicity (their elution order on reverse-phase HPLC).

(Hulmes et al., 1989). The TFA-treated filter was then removed from the sequencer, placed in a 15-mL polypropylene reaction tube, and exposed to more TFA vapors via the cartridge inlet line for 10 min. The tube was capped and incubated in an oven at 45 °C for 4 days. Similarly, cyanogen bromide cleavage was carried out on a filter placed on the side of a culture tube (13 × 100 mm) that contained 4 mg of CNBr in 0.4 mL of 70% TFA. The filter was wetted with 30 μL of the same TFA/CNBr solution. The tube was covered with parafilm and placed at room temperature in the dark for 24 h.

Enzyme-Catalyzed Cleavages. Lysyl bonds were hydrolyzed in 15 nmol of CM-cGS-PDE dissolved in 2 M urea/50 mM Tris-HCl, pH 9.0, using *Achromobacter* protease I (Masaki et al., 1981) at a protease:substrate ratio of 1:100 (w/w). After incubation (20 h, 37 °C), digestion was stopped by acidification with formic acid. Similar subdigests of smaller peptides used a protease:substrate ratio of 1:50 w/w for 11 h.

Ten nanomoles of reduced and S-pyridylethylated cGS-PDE was cleaved at arginyl bonds with TPCK-trypsin after citraconylation of lysyl residues. Citraconylation was carried out at pH 8.8 with a pH titrator by the addition in small aliquots of citraconic anhydride (32 μL) over 1 h at room temperature to the 6 M guanidine hydrochloride solution of protein. The modified protein was then extensively dialyzed against 50 mM ammonium bicarbonate at pH 8.5 and treated with 2.2% (w/w) TPCK-trypsin in two equal aliquots for 4 h at 37 °C. Following the tryptic digest, the citraconyl groups were removed by lowering the pH to 2.5 with 9% formic acid and incubating for 6 h at 37 °C.

Two large cyanogen bromide fragments, M2 (2.0 nmol) and M4 (2.3 nmol), were cleaved specifically at the amino end of aspartyl residues using *Pseudomonas fragi* endoproteinase Asp-N. Peptide was dissolved in 2 M urea, 25 mM methylamine, and 100 mM sodium phosphate, pH 8.0. The protease was added in two equal aliquots to give a final protease:substrate ratio of 1:75 (w/w), and the mixture was incubated at 37 °C for 28 h. Similarly, glutamyl bonds were cleaved with *Staphylococcal aureus* V8 protease (1:25, w/w) in 2 M urea/100 mM ammonium bicarbonate, pH 8.0 at 37 °C, for 33 h.

Subdigestion with chymotrypsin was carried out on a mixture of 300 pmol of arginyl peptide R10, 60 pmol of R4, and 50 pmol of R12 in 100 mM ammonium bicarbonate, pH 8.0, for 8 h at 37 °C using a protease:substrate ratio of 1:50 (w/w) and 10 mM *p*-aminobenzamidine to suppress trypsin. Products were separated on a C8 column. Subdigestion of peptide K1 with *Staphylococcal aureus* V8 protease used a protease:substrate ratio of 1:10 (w/w) at 37 °C for 24 h.

Peptide Purification. Mixtures of peptides from digests of the intact protein were initially separated by size-exclusion (Figure 1) on a tandem pair of either TSK-3000 SW(Bio-Rad) or TSK-2000 SW(LKB) columns equilibrated in 6 M guanidine hydrochloride and 10 mM sodium phosphate, pH 6.0 (Titani, 1986). A pooled fraction of large peptides was desalted and further purified by reverse-phase HPLC on an Altex Ultrapore RPSC-C3 column (4.6 × 75 mm). Peptides of moderate and smaller size were separated on a Synchropak RP-C18 column (4.1 × 250 mm) or on the microbore columns Aquapore RP-300-C8 (2.1 × 100 mm, 7 μm; Pierce) or RP-18 (2.1 × 30 mm, 5 μm; Brownlee Labs). Two HPLC systems were used in this study: a Varian 5000 liquid chromatograph and a Hewlett-Packard 1090 liquid chromatograph equipped with a diode array detector. Peptides were eluted with a linear gradient of 0.1% TFA/acetonitrile.

Automated Analyses. Amino acid analyses were determined with a Waters Picotag system as described by Bidlingmeyer et al. (1984). In the early phases of this study, peptides from the cyanogen bromide cleavage and the lysyl digest were sequenced with a Beckman 890 C sequencer as described previously (Charbonneau et al., 1985). An Applied-Biosystems Model 470A gas-phase sequencer equipped with an on-line Model 120A Pth analyzer was used during the later phases of this study.

Analysis of Structural Relationships. The search routine of the GENEPRO software (version 4.2; Riverside Scientific, Seattle) was utilized to identify homologous sequences within the PIR protein database (release 21.0, National Biomedical Research Foundation) and the GenBank (release 61.0) nucleotide database. Alignment of sequence pairs was performed on a VAX/VMS computer using the ALIGN program from the National Biomedical Research Foundation (Dayhoff et al., 1983). All alignments were scored by using the mutation data matrix and a gap penalty of 10. Optimal alignment of multiple PDE sequences was achieved by using the CLUSTAL programs described by Higgins and Sharp (1988). The dot matrix plotting routine of GENEPRO and the RELATE program from the National Biomedical Research Foundation were used to analyze internal homology.

Mass Spectrometry. Identification of the N-terminal blocking group and confirmation of the sequence of residues 1–10 were obtained by determining the mass of peptide K1-E1a. Mass spectra were obtained by Patrick Griffin (Genentech, Inc.) using a Bio-Ion time of flight plasma desorption mass spectrometer.

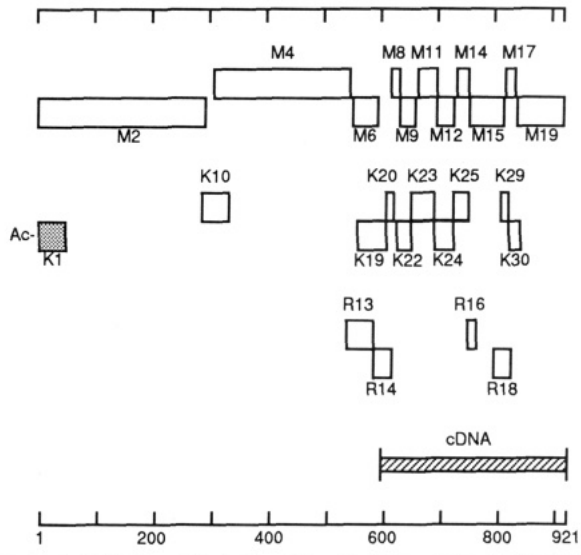

FIGURE 2: Relative orientation of major fragments used to display the cGS-PDE sequence. Prefixes M, K, and R denote peptides generated by cleavage at methionyl, lysyl, and arginyl bonds, respectively. The hatched bar identifies the N-terminal blocked peptide. The alignment of the partial cDNA sequence is indicated at the bottom. A detailed summary of the sequence analysis of these segments and peptides is illustrated in Figure 3.

Partial cDNA Clone. A detailed procedure for cloning and sequencing the cGS-PDE cDNA will be published elsewhere (W. K. Sonnenburg, P. J. Mullaney, and J. A. Beavo, unpublished results). In brief, oligonucleotides complementary to nucleic acids encoding the peptide sequences Glu-Met-Met-Met-Tyr-His-Met-Lys (residues 549–556) and Tyr-His-Asn-Trp-Met-His-Ala-Phe (residues 635–642) were synthesized and end-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase. These oligomers were used as hybridization probes for screening an unamplified bovine adrenal cortex cDNA library constructed λ ZAP II vector DNA (Stratagene). Putative clones were plaque-purified, subcloned by in vivo excision, and sequenced by a modification of the Sanger method (Biggins et al., 1983; Sanger et al., 1977). A single 2.2-kb cDNA clone encoding the cGS-PDE (residues 592–921) was identified and sequenced by constructing a set of nested deletions.

RESULTS

Overall Strategy. The general approach to sequence analysis of this 921-residue protein is illustrated in Figure 2. A set of fragments generated by cleavage at methionyl residues (peptides M2 through M19) was purified, analyzed, and aligned with overlapping fragments resulting from cleavage at lysyl residues (peptides K1 through K30). One additional peptide (R13) derived by arginine cleavage provided a crucial overlap in the middle of the molecule. Each peptide, or subdigestion fragment thereof, was subjected to Edman degradation. The results in Figure 3 provide a detailed proof of the unique sequence of cGS-PDE. A partial cDNA clone corresponding to the C-terminal residues 592–921 confirmed the alignments in the C-terminal third of the molecule and placed eight additional residues.

The digests at methionyl and at lysyl residues were each carried out with about 17 nmol of reduced and carboxymethylated protein whereas the arginyl digest utilized 10 nmol of reduced and pyridylethylated protein. In each case, the whole digest was first fractionated by size on TSK columns; individual proteins were then purified by reverse-phase HPLC of pooled fractions. Fragments M2 and M4 were each as large

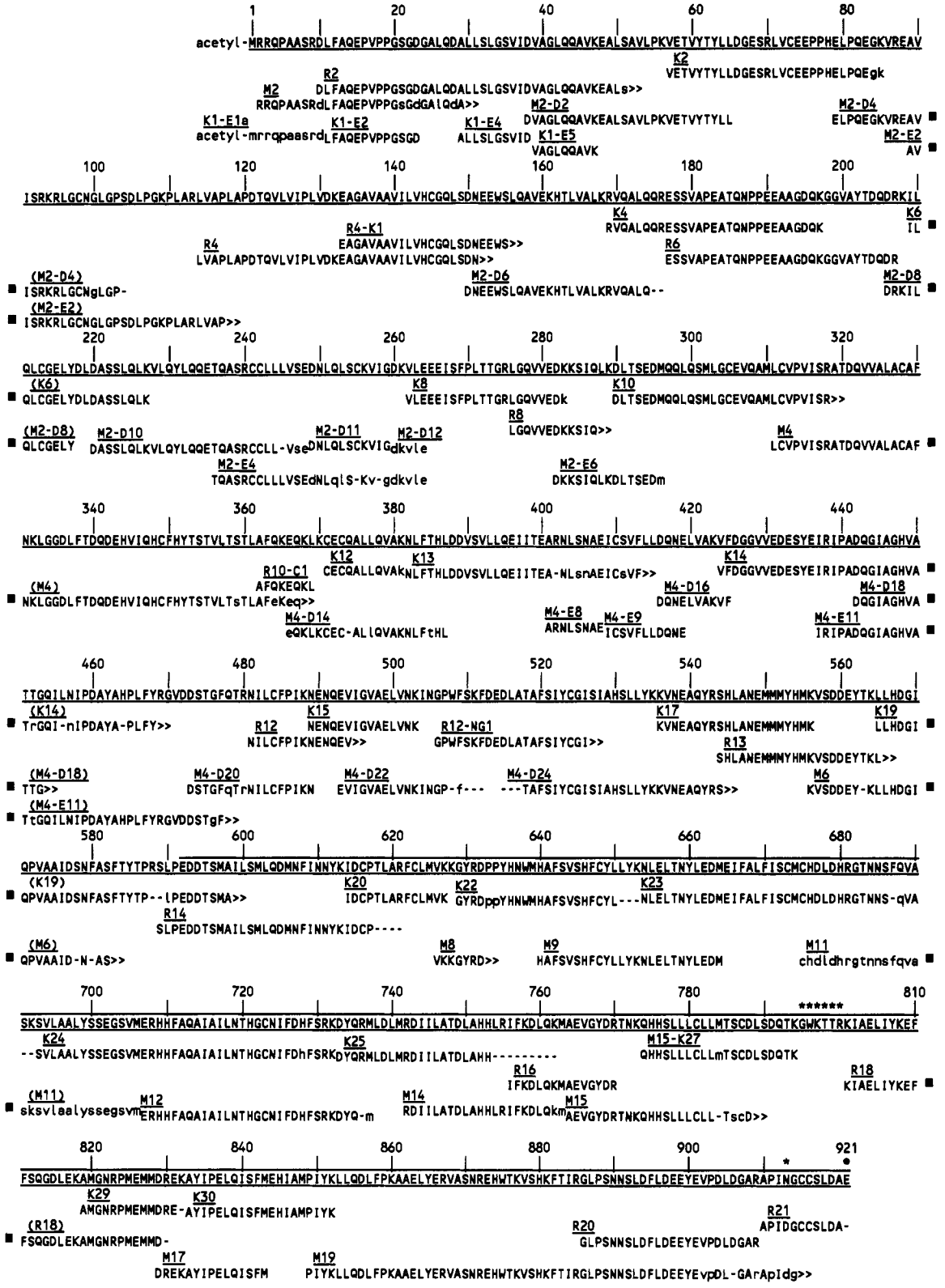

FIGURE 3: Detailed summary of the proof of the cGS-PDE sequence. The proven sequences of specific peptides, with underlined names, are given in one-letter code below the summary sequence. The names of the peptides use prefixes K, R, and M to designate peptides generated by cleavage at lysyl, arginyl, and methionyl bonds, respectively. Subpeptides have hyphenated suffixes of E, D, K, C, and NG indicating products of cleavage with *Staphylococcal aureus* V8 protease, *Pseudomonas fragi* endoproteinase Asp-N, *Achromobacter protease I*, chymotrypsin, and chemical cleavage with hydroxylamine, respectively. Peptide sequences in upper case letters were proven by Edman degradation; those in lower case letters were tentative identifications (due to a low yield near the end of a degradation or to uncertainty in a mixture of two peptides). Only the mass of N-terminal peptide K1-E1a was determined. A dash indicates a position where no identification was made; (≫) indicates a large peptide where the C-terminal sequence was not obtained; (■) at the right end of a line indicates that the sequence of a peptide continues below on another line, designated with a (■) at the left. The line above residues 592–921 identifies the portion also sequenced via its cDNA. The asterisk denotes a residue where the sequence was derived only from the cDNA.

as many intact proteins; each was subdigested at aspartyl or glutamyl residues to generate peptides of a size suitable for Edman degradation. All of the data fit the scheme in Figure 3, but only peptides necessary for the proof of structure are illustrated. The sequence of the N-terminal half of the molecule was largely provided by the complementarity of the arginyl cleavage products and peptides derived from M2 and M4. The sequence of the C-terminal half was largely provided by overlapping information among the lysyl peptides and the cyanogen bromide fragments.

Cleavage at Lysyl Residues. Digestion at 48 lysyl residues with *Achromobacter* protease I yielded a complex mixture (Figure 1). The Edman degradation of 18 lysyl peptides identified 466 residues of unique sequence as illustrated in Figure 3. One additional peptide (K1) was blocked; it was ultimately determined to be N^{α}-acetylated (see below) and to represent the N-terminal 47 residues of the protein. As with other digests, other peptides that did not contribute to the proof of structure are not illustrated in Figure 3.

Cleavage at Arginyl Residues. A set of peptides complementary to the lysyl cleavage products was derived by digesting N^{ϵ}-citraconylated, S-pyridylethylated protein with trypsin. From the digestion products, 12 purified peptides (R2 through R21 in Figure 3) provided unique sequences of 299 residues, some of which overlapped lysyl cleavage peptides, for example, in the region of residues 481–617. Peptides R13 and R18 (residues 544–588 and 801–830) were particularly useful in providing overlaps through the otherwise difficult sequence EMMMYHMKV and in the region of residues 801–820. Another peptide, R21, was recognized as the C-terminal peptide on the basis of its unique lack of arginine. Its sequence ended in LDA; analysis of the cDNA clone indicated that this was followed by Glu and a stop codon, confirming the location of R21 at the C-terminus.

Cleavage at Methionyl Residues. Cleavage with cyanogen bromide at 26 methionyl residues yielded the large fragments M2 and M4 comprising residues 2–296 and 311–550, as well as many smaller peptides, of which 9 were recovered in good yield after reverse-phase HPLC. The small peptides proved to be very useful as overlaps of the lysyl peptides in the C-terminal 355 residues. In fact, the sequence of residues 556–901 is almost entirely provided by the complementarity of peptides M6 through M19 with peptides K19 through K30. Other necessary information was provided by peptides R14 (residues 589–621) and R18 (residues 801–830) and by the cDNA clone (residues 795–800).

Sequence of the N-Terminal 295 Residues. Determination of the sequence of this segment was complicated by the observation that the N-terminus was blocked, as was K1, a 47-residue peptide derived from the N-terminus by digestion at lysyl residues. Virtually all of this segment corresponds to fragment M2 (residues 2–296). The N-terminal analysis of M2 overlapped R2, and these peptides were linked to other lysyl or arginyl cleavage products via analysis of overlapping peptides derived from M2 by digestion at aspartyl and glutamyl residues.

The blocked peptide K1 was assumed to be derived from the blocked N-terminus of the protein and, since it contained methionine, to overlap M2. To verify this, 2 nmol of a mixture of large lysyl peptides containing K1 was subjected to prolonged treatment with V8 protease, and the products were separated on a C8 column. Three of the purified peptides corresponded to residues 11–23, 29–38, and 39–47, respectively, indicating that aspartyl bonds had been cleaved and verifying that K1 overlapped M2. The two most hydrophilic peptides, K1-E1a and K1-E1b, were both blocked and appeared to have similar compositions. Treatment of each peptide with gaseous TFA for 4 days on the glass fiber filter used for its Edman degradation exposed the sequence Ser-Arg, corresponding to residues 8 and 9 in Figure 3. After this acid treatment, the two filters were treated together in a sealed tube with CNBr (10 mg/mL in 70% TFA) for 24 h and then subjected a third time to Edman degradation. This yielded the single sequence RRQPAA, which is identical with the N-terminus of M2.

Aliquots of K1-E1a and K1-E1b were mixed and analyzed together by time-of-flight plasma desorption mass spectrometry. The results indicated two principal components, of mass $(M+H)^+ = 1230.3$ and 1246.9, respectively. The smaller corresponded to the mass of N^{α}-acetyl-MRRQPAASRD; the larger had an additional oxygen atom and corresponded to the methionine sulfoxide form of the same peptide.

The remainder of the proof of sequence of the first 295 residues is evident from the summary in Figure 3. It appeared to be rigorous except between Gly-259 and Val-262 where the proof was based on tentative identification within M2-D12 and M2-E4. Analysis of M2-D12 involved the interpretation of a mixture in which it was the major peptide; analysis of M2-E4 in this region was of low yield at the end of its Edman degradation.

Residues 296–555. Most of this segment was derived from peptides generated by subdigestion of the large CNBr fragment M4 (residues 311–550) and their overlaps with lysyl and arginyl cleavage products from the whole protein. A critical overlap was also provided by cleavage with hydroxylamine of an Asn–Gly bond in R12 to generate the sequence beginning at Gly-506 in R12-NG1. At the N-terminal end of this segment, peptide K10 spanned the sequence between M2 and M4. At the C-terminal end, K17 and R13 overlapped the C-terminus of M4 and the N-terminus of M6 through a difficult six-residue segment containing four methionines (residues 550–555). At residue 478, both Gln and Arg were observed, but Gln appeared to be the predominant residue; neither identification was confirmed in an overlapping structure. The appearance of both residues in a single cycle of the Edman degradation suggests microheterogeneity which might be explained by the presence of multiple alleles.

Residues 556–921. Overlapping lysyl, arginyl, and methionyl cleavage products provided this segment of sequence, but five regions presented difficulties. First, arginine-588 was not observed during Edman degradation but inferred from the cleavage that yielded R14. Second, peptide M11 was sequenced only as a minor component in a mixture (1:3) with M12; thus, the identification of residues 687–688 and 691–692 was somewhat tentative until confirmed by the cDNA sequence. Residues 795–800 were not recovered in the digests examined, and the sequence (GWKTTR) is derived solely from the cDNA. Residue 913 was observed to be aspartic acid in R21 and M19 but asparagine according to the cDNA; we concluded that the native protein has asparagine at this locus because it is known that aparaginyl-glycine sequences rearrange spontaneously to yield products that include Asp-Gly and other structures that prevent Edman degradation (Bornstein & Balian, 1977; Takio et al., 1986). Finally, as already described, the C-terminal glutamyl residue was identified by the cDNA.

This proof of sequence depends primarily upon Edman degradations; of the 921 residues in cGS-PDE, 913 were identified directly by analysis of peptides. The nature of the proof of structure involves many redundant observations that

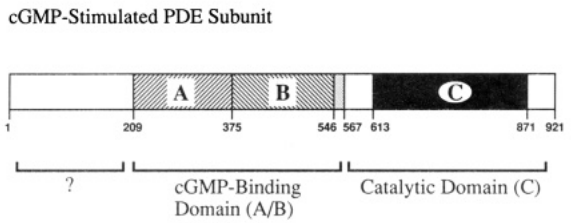

FIGURE 4: Domain organization of the cGMP-stimulated PDE subunit. Conserved sequences within the catalytic domain (C) are shown as a solid black box, whereas boxes containing diagonal lines designate the two internally homologous subdomains (A and B) of the conserved cGMP binding domain. Residue numbers, given below, denote the boundaries of conserved segments. The rationale for the assignment of precise boundary positions is given under Results.

increase the reliability of the conclusions. For example, in the N-terminal 100 residues, about 70% were sequenced in more than 1 overlapping peptide, and the N-terminus itself was confirmed by mass spectrometry. In the latter stages of the investigation, a partial cDNA clone became available so that in the C-terminal 330 residues (592–921) the proof was 3-fold at many loci with an accumulation of evidence derived by sequencing both the protein and the cDNA. The sequence of the cDNA clone also provided eight residue identifications (795–800, 913, and 921). Nonetheless, certain parts of the proof are less certain, and these are listed below. At residues 206 and 424, peptides overlap by only two residues. At residues 260–264, an overlap linking M2-D11 to K8 depends upon two tentative sequences, one at the end of an Edman degradation and the other in an impure peptide (M2-D12). Residues 359, 478, and 480 were tentatively identified in only one sequence analysis.

Search for Homologous Proteins. A search of the PIR and GenBank databases revealed no sequences other than PDEs that display significant similarity to the cGS-PDE. With the exception of PDE 1 from *S. cerevesiae* (Nikawa et al., 1987) and secreted PDE from *Dictyostelium* (Lacombe et al., 1986), all PDE sequences including the cGS-PDE have a conserved segment of about 250 residues (Charbonneau et al., 1986). The conserved region (segment C, Figure 4) from the cGS-PDE includes residues 613–871 and shows 28–35% sequence identity to corresponding regions of nine other PDEs (Table I).

The structural relationship between the cGS-PDE and the α and α' catalytic subunits of the rod and cone photoreceptor PDEs is not restricted to segment C, but also includes an additional region of homology (both segments A and B, Figure 4) that is located within the N-terminal part of the molecule (Charbonneau et al., 1990). The extensive nature of the structural relationship among these 3 PDEs is illustrated by the fact that all but the first 169 residues of the cGS-PDE can be aligned with the *complete* α and α' subunit sequences to give overall sequence identities of 28 and 29%, respectively. Three regions of the cGS-PDE having significantly reduced sequence similarity to the photoreceptor isozymes serve to delimit boundaries of the two conserved regions (Figure 4). Segment C (Figure 4) and segment A/B (Figure 4) of the cGS-PDE are separated by a region of about 45 residues that shows only 9% sequence identity with the α subunit and requires a 21-residue gap for optimal alignment. In addition, among the photoreceptor PDE subunits and the cGS-PDEs, segment C is more conserved than segment A/B. The C-terminal boundary of segment C is established by a 49-residue region having only 14% sequence identity with the α subunit. Similarly, a region of 56 residues with 16% identity to the α subunit delimits the N-terminal boundary of segment A/B.

Table I: Homology between the Catalytic Domains of the cGMP-Stimulated PDE and Other PDE Isozymes

		cGMP-stimulated PDE (613–868)	
phosphodiesterase[a]	catalytic domain residues	% sequence identity	alignment score[b]
61-kDa CaM	193–446	28	16.0
Drosophila dunce	223–489	33	16.7
rat dnc-1	161–427	31	16.9
rat DPD	163–429	31	18.4
rat PDE 1	*c*	33	18.4
rat PDE 3	214–480	31	18.1
ROS α subunit	535–803	35	27.0
cone α' subunit	535–802	35	24.6
S. cerevisiae PDE 2	241–478	23	5.0

[a] Sequences compared to the cGS-PDE were taken from the following sources: 61-kDa calmodulin-dependent PDE (Charbonneau et al., manuscript submitted); *Drosophila* dunce PDE (Chen et al., 1986); rat dnc-1 (Davis et al., 1989); rat DPD (Colicelli, 1989) rat PDE1 (Swinnen et al., 1989a); rat PDE 3 (Swinnen et al., 1989b); ROS PDE α subunit (Ovchinnikov et al., 1987); cone α' subunit (Li et al., 1990); PDE2 from *S. cerevisiae* (Sass et al., 1986). [b] Alignments and alignment scores were generated by using the ALIGN program (Dayhoff et al., 1983) suplied by the National Biomedical Research Foundation. [c] Residue numbers are not given for rat PDE 1 since the complete sequence has not been reported; sequences aligned were taken from Figure 4 of Swinnen et al. (1989b)

Optimization of the alignment between the photoreceptor and cGS-PDEs indicates that the boundaries of the conserved segment A/B (Figure 4) reported earlier in Figure 5 of Charbonneau et al. (1990) can be expanded to include about 50 additional residues at the N-terminus. Thus, residues 209–567 of the cGS-PDE encompass segment A/B and can be aligned with corresponding sequences of the α subunit (residues 60–469) and α' subunit (residues 57–467) to give 28 and 30% sequence identity and alignment scores (Dayhoff et al., 1983) of 13 standard deviations (SD) and 17 SD, respectively. This revision in domain boundaries is consistent with the multiple alignment of all three cGMP binding PDEs as discussed above and the position of internally homologous segments within the domain (see below). Although these methods provide statistical data consistent with the stated boundaries, it should be emphasized that precise boundaries of the proposed domains cannot be rigorously established in proteins having this level of overall sequence similarity.

Internal Homology. The conserved, segment A/B of the cGS-PDE, and the photoreceptor PDE subunits possess two internally homologous repeats (A and B, Figure 4). In our original report (Charbonneau et al., 1990), we indicated that the 2 internal repeats were separated by a 61-residue segment that was unrelated to either repeat but similar among all three PDEs. An examination of internal homology within the cGS-PDE indicates that the two internal repeats A and B are contiguous (Figure 4). Their alignment in Figure 5 demonstrates the homologous relationship among the repeats within the cGS-PDE and rod photoreceptor α subunit. The repeats of the cGS-PDE (Figure 5) show 23% sequence identity with a Dayhoff alignment score of 9.0 (Table II). The internal repeats of the α and α' subunits are also positioned in tandem. However, attempts to align analogous repeats within each photoreceptor PDE subunit shows that the C-terminal repeat (corresponding to segment B in Figure 4) is longer and contains a 25-residue sequence that is not similar to any sequences within the N-terminal repeat or either repeat of the cGS-PDE. Apparently, the C-terminal repeat of the photoreceptor PDEs, but not the cGS-PDE, acquired a 25-residue insert during evolution. If this insert is excluded, the tandem repeats of the

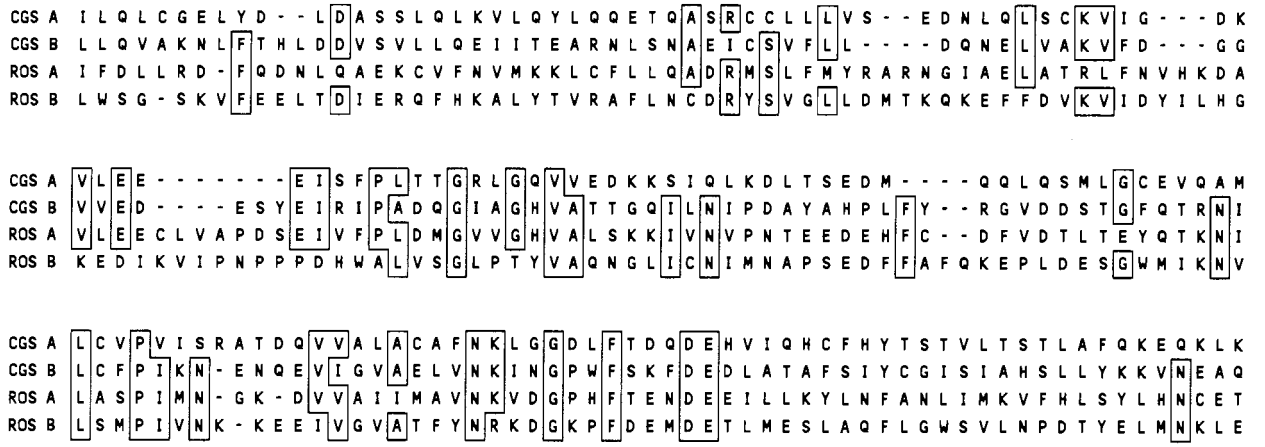

FIGURE 5: Alignment of the two internal repeats from the cGMP-stimulated PDE (cGS-PDE) and the α subunit (Ovchinnikov et al., 1987) of the photoreceptor PDE (ROS) from rod outer segments. Sequences are given in single-letter code; hyphens denote gaps required for optimal alignment. Boxes enclose residues that are identical in at least three of the four sequences. The following sequences were aligned: CGS A (residues 209–370); CGS B (residues 376–541); ROS A (60–234); and ROS B (241–290; 316–443). The alignment shown was obtained with the CLUSTAL program for multiple sequence alignment (Higgins & Sharp, 1988).

Table II: Internal Homology[a] in cGS-PDE and the α and α' Subunits of the Photoreceptor PDEs

PDE segment[b]	cGS A	cGS B	ROS α A	ROS α B[c]	cone α' A	cone α' B[c]
cGS A						
cGS B	9.0 (23)					
ROS α A	7.1 (27)	9.7 (28)				
ROS α B	3.9 (18)	9.4 (28)	7.0 (21)			
cone α' A	7.4 (25)	8.4 (32)	27.6 (48)	9.4 (22)		
cone α' B	5.3 (20)	10.9 (30)	6.3 (22)	43.0 (77)	7.7 (25)	

[a] Similarities among internally homologous segments (A and B) are given as alignment scores produced by the ALIGN program (Dayhoff et al., 1983) or as percent sequence similarities, in parentheses. [b] Sequences of the α and α' subunits of the rod (ROS) and cone PDEs were taken from Ovchinnikov et al. (1987) and Li et al. (1990), respectively. Segments A and B for each enzyme are defined as follows: cGS A (209–375), cGS B (376–546), ROS α A (60–240), ROS α B (241–290, 316–450), cone α' A (60–238), and cone α' B (239–288, 314–448). [c] Note that for these alignments segment B from the α and α' subunits does not include residues 291–315 and 289–313, respectively, for reasons justified in the text (see Results).

α and α' subunits show 21% and 25% sequence identity, respectively, and give statistically significant alignment scores (Table II).

Discussion

The subunit molecular weight of cGS-PDE is calculated to be 103 244 including the N-terminal acetyl group. This is in good agreement with earlier estimates of 102 000–107 000 by SDS–PAGE and of 201 000 and 240 000 for the native dimer by density gradient centrifugation and gradient electrophoresis, respectively (Martins et al., 1982; Yamamoto et al., 1983). It is clear from the present analysis that the dimeric molecule must be composed of identical chains. Three Asn-Xaa-Ser/Thr sequences (at Asn-403, -684, and 890) that code for glycosylation sites in extracellular proteins are not glycosylated in cGS-PDE. Aside from the *N*-acetyl blocking group, no additional posttranslational modifications were detected.

Domain Organization of the cGS-PDE. The organization of conserved segments within the cGS-PDE suggests that it is comprised of at least three functional domains (Figure 4). A 259-residue conserved segment (C in Figure 4) near the C-terminus of the PDE is observed in all PDE sequences of metazoan origin (Table I). The limited proteolysis and direct photolabeling studies of Stroop et al. (1989) provide evidence that a 36-kDa fragment encompassing this conserved region is active. Moreover, a fully active 36-kDa fragment from the 61 kDa CaM-PDE (Charbonneau et al., manuscript submitted) contains the analogous, conserved segment. Thus, data for PDEs from two different isozyme families show that the conserved domains common to most PDEs contain the catalytic site.

A second domain of the cGS-PDE is delimited by sequences (A and B, Figure 4) which are homologous to only two other PDE sequences, within the α and α' subunits of the photoreceptor PDE (Charbonneau et al., 1990). All three of these PDEs have in common the ability to bind cGMP at specific noncatalytic sites (Erneux et al., 1981; Miot et al., 1985; Gillespie & Beavo, 1988, 1989), suggesting that these conserved, N-terminal features might delineate the location of cGMP binding site(s). Using [^{32}P]cGMP, Stroop et al. (1989) demonstrated the selective photolabeling of cGMP binding sites having properties expected for the noncatalytic regulatory sites of the enzyme (Erneux et al., 1981; Miot et al., 1985). The N-terminal conserved region was contained within several different fragments labeled under these conditions and was assigned a function in the formation of allosteric, cGMP binding site(s) (Stroop et al., 1989; Charbonneau et al., 1990).

The conserved cGMP binding domain is comprised of two internally homologous repeats. The role of each internal repeat in the formation of the cGMP binding site(s) poses an interesting question. A reevaluation of the binding properties of the cGS-PDE using [^{3}H]cGMP and different binding conditions (Stroop and Beavo, submitted for publication) indicates that there are *two* noncatalytic cGMP binding sites per homodimer rather than the single mole per dimer reported previously (Martins et al., 1982). Assuming there is only one noncatalytic cGMP site per monomer, the significance of the repeating pattern in each subunit is unclear. The cGMP binding site might be formed as a cleft or pocket between interacting internal repeats in each monomer where an element of symmetry would result from their homologous character. Alternatively, each cGMP may bind at a site formed at an

interface between the two subunits of the holoenzyme, as suggested by Stroop et al. (1989). In this case, two repeats would come together to form each cGMP binding site, but the repeats would reside on separate polypeptide chains. Further evaluation of the possible arrangements will require precise localization of the cGMP binding site using site-specific labeling or mutational analysis of expressed cDNA clones.

As noted under Results, sequence alignments indicate that the C-terminal repeat (corresponding to segment B, Figure 4) of the photoreceptor PDE subunits possess a 25-residue insert that is absent from the cGS-PDE. The functional significance of this insert is not known. However, it may contribute to the greatly enhanced cGMP affinity (≥100-fold) of the photoreceptor PDE sites over those of the cGS-PDE.

Several observations indicate that a third domain of as yet unknown function may exist at the N-terminus of the cGS-PDE. About 200 N-terminal residues are not included within the conserved, cGMP binding region, and nearly 150 of these residues cannot be aligned with the photoreceptor catalytic subunits (Figure 4). Since these residues have no counterpart within the photoreceptor forms, it is unlikely that they have an essential role in cGMP binding. This domain may be involved in localizing the enzyme to a specific intracellular compartment or may be a site of interaction with other regulatory molecules. A search of the PIR protein sequence database, conducted with the sequence of the N-terminal domain (residues 1–210) or the complete sequence, revealed no proteins having significant similarity with this domain. One interesting feature of this region of the molecule is its relatively high content of Pro residues. The mole percent of Pro in the first 150 residues of the N-terminal domain is 3 times higher than that of the remainder of the molecule (residues 211–921).

Possible Evolutionary Mechanisms. The cGS-PDE belongs to a family of evolutionarily related PDEs that all possess a conserved catalytic domain. However, the cGS-PDE is not structurally related to either PDE 1 from *S. cerevisiae* (Nikawa et al., 1987) or a PDE from *Dictyostelium* (Lacombe et al., 1986). In addition, the cGS-PDE shows no significant sequence similarity to another protein family that binds cyclic nucleotides, which includes the regulatory subunits of protein kinases and the catabolite activator protein of *Escherichia coli* (Takio et al., 1984). Apparently, the cyclic nucleotide binding sites of these three classes of proteins have distinct evolutionary origins or diverged so long ago and at such a rate that their common ancestry is no longer detectable. If the latter were the case, only three-dimensional analysis could detect a structural relationship.

The conserved catalytic and regulatory domains share a common function, the ability to bind cyclic nucleotides, and it is possible that the two sites evolved by duplication of an ancestral cyclic nucleotide binding domain. If so, the catalytic domain of the PDEs should display tandem repeats similar to those within the cGMP binding domain of the cGS-PDE. Internal repeats with sequence identities ranging from 16 to 22% have been observed in catalytic domains of the 61-kDa calmodulin-dependent PDE from bovine brain (Charbonneau et al., manuscript submitted), *Drosophila* dunce PDE (Chen et al., 1986), rat dnc-1 PDE (Davis et al., 1989), and the cGS-PDE. In each case, the repeats are separated by a single segment that varies in length and has little similarity to corresponding regions of other isozymes [see Charbonneau et al. (1986) or Swinnen et al. (1989a)]. The strongest evidence of internal homology comes from the calmodulin-dependent PDE where interna repeats of about 120 residues can be aligned to give 16% sequence identity (32% including chemically conserved residues) and an alignment score of 4.1 SD (Dayhoff et al., 1983). Two segments (residues 625–731 and 763–867) of the cGS-PDE catalytic domain display 22% sequence identity but give a marginal alignment score of 2.4 SD. If the two sites are related to a common ancestor, it is not clear why the similarity between internal repeats of the catalytic domains is less than that of the regulatory domains. Perhaps, acquisition of the ability to hydrolyze the phosphodiester bond required a concomitant reduction in symmetry between the internal domains.

The strongest evidence that the catalytic and regulatory domains of the cGS-PDE are related by an ancient gene duplication event is provided by the homology between ≈90-residue segments in each domain. Residues 650–740 from the N-terminal half of the catalytic domain and residues 376–458 from segment B (Figure 4) of the cGMP binding domain have 22% sequence identity with an alignment score (Dayhoff et al., 1983) of 5.1 SD. No convincing evidence of homology was apparent when other segments of the two domains were compared. Although the evidence for a structural relationship between the domains of the cGS-PDE is marginal, these data are consistent with an evolutionary process involving gene duplication. Following duplication, it is assumed that the two domains diverged extensively as each one acquired the structural features unique to its specialized function, i.e., allosteric regulation or hydrolysis of the phosphodiester bond. Thus, the somewhat limited and relatively low degree of sequence similarity between the two domains may simply reflect their functional specialization.

Many multidomain proteins are chimeric in character, e.g., cGMP-dependent protein kinase [see Takio et al. (1984)], and have probably evolved by ancestral gene splicing events which resulted in the fusion of two domains with evolutionarily distinct origins. In contrast, the evidence for internally homologous segments, noted above, suggests that the cGS-PDE acquired its cyclic nucleotide binding domains through mechanisms involving tandem gene duplication rather than gene fusion.

ACKNOWLEDGMENTS

We thank Dr. Patrick Griffin for performing the mass spectrometric analysis of the blocked N-terminal peptide and Roger Wade, Santosh Kumar, and Maria Harrylock for their assistance with amino acid analyses and gas-phase sequencing.

REFERENCES

Beavo, J. A. (1988) *Adv. Second Messenger Phosphoprotein Res. 22*, 1–38.

Beavo, J. A., & Reifsynder, D. H. (1990) *Trends Pharmacol. 11*, 150–155.

Beavo, J. A., Hardman, J. G., & Sutherland, E. W. (1971) *J. Biol. Chem. 246*, 3841–3846.

Bidlingmeyer, B. A., Cohen, S. A., & Tarvin, T. L. (1984) *J. Chromatogr. 336*, 93–104.

Biggins, M. D., Gibson, T. J., & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3963–3965.

Bornstein, P., & Balian, G. (1977) *Methods Enzymol. 47*, 132–145.

Charbonneau, H., Walsh, K. A., McCann, R. O., Prendergast, F. G., Cormier, M. J., & Vanaman, T. C. (1985) *Biochemistry 24*, 6762–6771.

Charbonneau, H., Beier, N., Walsh, K. A., & Beavo, J. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 9308–9312.

Charbonneau, H., Prusti, R. K., Le Trong, H., Sonnenburg, W. K., Mullaney, P. J., Walsh, K. A., & Beavo, J. A. (1990)

Proc. Natl. Acad. Sci. U.S.A. 87, 288–292.
Chen, C. N., Denome, S., & Davis, R. L. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 93139–9317.
Colicelli, J., Birchmeier, C., Michaeli, T., O'Neill, K., Riggs, M., & Wigler, M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 3599–3603.
Davis, R. L., Takayasu, H., Eberwine, M., & Myres, J. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 3604–3608.
Dayhoff, M. O., Barker, W. C., & Hunt, L. T. (1983) *Methods Enzymol. 91*, 524–545.
Erneux, C., Couchie, D., Dumont, J. E., Baraniak, J., Stec, W. J., Garcia Abbad, E., Petridis, G., & Jastorff, B. (1981) *Eur. J. Biochem. 115*, 503–510.
Gillespie, P. G., & Beavo, J. A. (1988) *J. Biol. Chem. 263*, 8133–8141.
Gillespie, P. G., & Beavo, J. A. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4311–4315.
Gross, E. (1967) *Methods Enzymol. 11*, 238–255.
Higgins, D. C., & Sharp, P. M. (1988) *Gene 73*, 237–244.
Hulmes, J. D., Miedel, M. C., & Pan, Y. C. E. (1989) in *Techniques in Protein Chemistry* (Hugli, T. E., Ed.) pp 7–16, Academic Press, San Diego, CA.
Lacombe, M.-L., Podgorski, G., Franke, J., & Kessin, R. (1986) *J. Biol. Chem. 261*, 16811–16817.
Le Trong, H., Newlands, G. F., J., Miller, H. R. P., Charbonneau, H., Neurath, H., & Woodbury, R. G. (1989) *Biochemistry 28*, 391–395.
Li, T., Volpp, K., & Applebury, M. L. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 293–297.
Martins, T. J., Mumby, M. C., & Beavo, J. A. (1982) *J. Biol. Chem. 257*, 1973–1979.
Masaki, T., Tanabe, M., Nakamura, K., & Soejima, M. (1981) *Biochim. Biophys. Acta 660*, 44–50.
Miot, F., Van Haastert, P. J. M., & Erneux, C. (1985) *Eur. J. Biochem. 149*, 59–65.
Nikawa, J., Sass, P., & Wigler, M. (1987) *Mol. Cell. Biol. 7*, 3629–3636.
Ovchinnikov, Y. A., Gubanov, V. V., Khramtsov, N. V., Ischenko, K. A., Zagranichny, V. E., Muradov, K. G., Shuvaeva, T. M., & Lipkin, V. M. (1987) *FEBS Lett. 223*, 169–173.
Sanger, F., Miklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Sass, P., Field, J., Nikawa, J., Toda, T., & Wigler, M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 9303–9307.
Stroop, S. D., Charbonneau, H., & Beavo, J. A. (1989) *J. Biol. Chem. 264*, 13718–13725.
Swinnen, J. V., Joseph, D. R., & Conti, M. (1989a) *Proc. Natl. Acad. Sci. U.S.A. 86*, 5325–5329.
Swinnen, J. V., Joseph, D. R., & Conti, M. (1989b) *Proc. Natl. Acad. Sci. U.S.A. 86*, 8197–8201.
Takio, K., Wade, R. D., Smith, S. B., Krebs, E. G., Walsh, K. A., & Titani, K. (1984) *Biochemistry 23*, 4207–4218.
Takio, K., Blumenthal, D. K., Walsh, K. A., Titani, K., & Krebs, E. G. (1986) *Biochemistry 25*, 8049–8057.
Titani, K. (1986) in *Methods in Protein Sequence Analysis* (Walsh, K. A., Ed.) pp 171–187, Humana Press Inc., Clifton, NJ.
Wada, H., Manganiello, V. C., & Osborne, J. C. (1987) *J. Biol. Chem. 262*, 13938–13945.
Yamamoto, T., Manganiello, V. C., & Vaughan, M. (1983) *J. Biol. Chem. 258*, 12526–12533.