Photoaffinity Labeling with UMP of Lysine 992 of Carbamyl Phosphate Synthetase from *Escherichia coli* Allows Identification of the Binding Site for the Pyrimidine Inhibitor†

Javier Cervera,‡ Elena Bendala,‡ Hubert G. Britton,§ Jorge Bueso,‡ Zeina Nassif,|| Carol J. Lusty,|| and Vicente Rubio*,‡

Instituto de Investigaciones Citológicas (FIB and CSIC), Amadeo de Saboya, 4, 46010 Valencia, Spain, Department of Physiology, St. Mary's Hospital Medical School, Paddington, London, U.K., and Department of Molecular Genetics, The Public Health Research Institute, New York, New York 10016

Received October 26, 1995⊗

ABSTRACT: UMP is a highly specific reagent for photoaffinity labeling of the allosteric inhibitor site of carbamyl phosphate synthetase (CPS) from *Escherichia coli* and has been found to be photoincorporated in the COOH-terminal domain of the large subunit [Rubio *et al.* (1991) *Biochemistry 30*, 1068−1075]. In the present work we identify lysine 992 as the residue that is covalently attached to UMP. This identification is based on two lines of evidence. First, [^{14}C]UMP is found to be incorporated between residues 939 and 1006, as shown by peptide mapping and by mass estimates of [^{14}C]UMP-peptides generated by chemical and enzymatic cleavage of CPS. Secondly, we have purified two radioactive peptides derived exclusively from those enzyme molecules (∼5% of the total enzyme) that had incorporated [^{14}C]-UMP. Edman analyses show the sequences of the labeled peptides (989)LVNXVHEGRPHIQD and (989)-LVNXVHE to be overlapping. Since neither a phenylthiohydantoin (Pth) derivative (in cycle 4) nor any radioactivity is released from the membrane during sequencing, we can conclude that Lys992 and [^{14}C]-UMP form a covalent adduct that remains bound to the membrane. Formation of this adduct agrees with all of the evidence and with the finding that UMP labeling prevents trypsin cleavage at Lys992. Lysine 992 is invariant in those CPSs that are inhibited by UMP, and is located 30 residues upstream of the site whose phosphorylation in hamster CAD reduces inhibition of CAD by UTP. Multiple sequence alignment of the residues surrounding Lys992 of the *E. coli* enzyme and the corresponding residues of the yeast and animal enzymes supports the existence of a uridine nucleotide binding fold in this region of the protein. We conclude that sequence changes in the binding fold provide a structural basis for the different regulatory properties found among CPSs I, II, and III.

Carbamyl phosphate synthetase catalyzes the first committed step in the urea, arginine, and pyrimidine biosynthetic pathways. The enzyme from most sources is subject to allosteric regulation, the nature of the effectors depending on the physiological role of the synthetase. For example, the ureotelic enzyme is activated by acetylglutamate[1] (Hall *et al.*, 1958) whereas the *Escherichia coli* enzyme, which is involved in both arginine and pyrimidine synthesis, is inhibited by UMP and is activated by IMP and by ornithine (Meister, 1989).

With the preparation of crystals suitable for X-ray diffraction studies (Marina *et al.*, 1995; Thoden *et al.*, 1995), the three-dimensional structure of carbamyl phosphate synthetase may be determined soon, conferring particular value to the location of functional residues in the protein. Using photoaffinity labeling techniques, we have shown previously that the inhibitor UMP and the activator IMP of the *E. coli* enzyme (a heterodimer of two subunits of M_r 41 400 and 117 700) bind in the COOH-terminal domain of the large subunit, a domain of about 18 kDa (Rubio *et al.*, 1991; Bueso *et al.*, 1994). In earlier studies, Rodríguez-Aparicio *et al.* (1989) have shown that the activator acetylglutamate binds in the same region of ureotelic carbamyl phosphate synthetase. Based on these findings together with the reported domain structure of the enzyme (Powers-Lee & Corina, 1986; Guadalajara, 1987; Evans & Balon, 1988; Marshall & Fahien, 1988; Rubio *et al.*, 1991; Cervera *et al.*, 1993), we have proposed that the COOH-terminal region of carbamyl phosphate synthetase functions as a regulatory domain. Recent elegant experiments involving domain-swapping between the *E. coli* and the pyrimidine-specific hamster carbamyl phosphate synthetase (Liu *et al.*, 1994) confirm that the COOH-terminal domain of the *E. coli* enzyme binds the nucleotide effectors, and that the equivalent domain in

† Supported by Grants PB90-0665 and PB93-1004 from the Dirección General de Investigación Científica y Técnica (DGICYT), 95/1227 from the Fondo de Investigaciones Sanitarias (FISS), and GM25846 from the National Institutes of Health. Elena Bendala was a Severo Ochoa Fellow of the Ayuntamiento de Valencia, and Jorge Bueso is a Predoctoral Fellow of the Generalitat Valenciana.

* Address correspondence to this author at Instituto de Investigaciones Citológicas, Amadeo de Saboya, 4, 46010 Valencia, Spain. Phone: +34-6-3698500. FAX: +34-6-3601453. E-mail: rubio@ochoa.fib.es.

‡ Instituto de Investigaciones Citológicas.

§ St. Mary's Hospital Medical School.

|| The Public Health Research Institute.

⊗ Abstract published in *Advance ACS Abstracts,* May 15, 1996.

[1] Abbreviations: acetylglutamate, *N*-acetyl-L-glutamate; ACTH, adrenocorticotrophic hormone; BSA, bovine serum albumin; CNBr, cyanogen bromide; GnRH, gonadotrophin-releasing hormone; HPLC, high-performance liquid chromatography; PAGE, polyacrylamide gel electrophoresis; PMSF, phenylmethylsulfonyl fluoride; Pth, phenylthiohydantoin; SDS, sodium dodecyl sulfate; Tris, tris(hydroxymethyl)-aminomethane; V8 protease, V8 staphylococcal protease.

S0006-2960(95)02549-9 CCC: $12.00

the hamster enzyme binds its inhibitor, UTP, and its activator, phosphoribosyl pyrophosphate. Further evidence implicating the COOH-terminal domain in allosteric regulation comes from the finding of Carrey *et al.* (1985) that the phosphorylation of the pyrimidine-specific hamster enzyme at Ser1406, which is within the COOH-terminal domain, decreases the inhibition by UTP.

To further define the structure of the COOH-terminal domain and its function in the allosteric control of carbamyl phosphate synthetase, we have identified the site of UMP binding by sequence analysis of a unique ^{14}C-labeled peptide, isolated after photochemical addition of [^{14}C]UMP to the enzyme. We demonstrate that photolabeling is due to the formation of a covalent adduct between UMP and lysine 992. Analysis of the sequence around this lysine in the *E. coli* enzyme and in other carbamyl phosphate synthetases supports the existence of an uridine nucleotide binding fold located entirely within the COOH-terminal domain of the protein.

EXPERIMENTAL PROCEDURES

Chemicals and Enzymes. [U-^{14}C]UMP (450 mCi/mmol) from CEA (Centre d'Etudes Nucleaires, Saclay, France) or, when it became unavailable, [2-^{14}C]UMP (46.1 mCi/mmol; from Sigma) was used. ^{14}C-labeled standard proteins for SDS−PAGE were from Amersham (14.3−200 kDa) or Gibco-BRL (3−43 kDa). Proteases (sequencing grade) were from Boehringer Mannheim or Sigma. Somatostatin and GnRH were a gift from Serono. Bradykinin, oxytocin, and ACTH were from Boehringer Mannheim, Sandoz, and Ciba, respectively. Other protein and peptide markers were from Sigma. All reagents were of the highest quality available from Merck, Sigma, Boehringer Mannheim, or ICN.

E. coli carbamyl phosphate synthetase was purified and was photoaffinity labeled with [^{14}C]UMP as described (Rubio *et al.*, 1991). The 25.8-kDa COOH-terminal fragment of the large subunit was generated by limited V8 protease digestion and isolated by HPLC (Rubio *et al.*, 1991). The specific radioactivity of the fragment was (1.5−2) × 10^6 cpm/mg of protein.

Reversed-Phase HPLC. Peptides were separated on a Vydac 5-μm C-18 column (type 218TP54, 0.46 × 25 cm) using (unless stated) a 60-min linear gradient (0 to 76%) of acetonitrile containing 0.1% trifluoracetic acid, at a flow rate of 1 mL/min. Fractions were collected at 1-min intervals, and the radioactivity was determined by liquid scintillation counting. For sequence analysis, peak fractions detected by monitoring the absorbance at 214 nm were collected manually.

Estimation of the Mass of Labeled Fragments by Gel Filtration Chromatography. The method of Belew *et al.* (1978), which utilizes a Sephacryl S-200 column, was modified by using as eluant 0.1 M Tris-HCl, pH 7.2, containing 0.4% SDS, instead of 6 M guanidine hydrochloride. The sample, mixed with 3 mg of BSA, 2 mg of RNAase, 2.7 mg of GnRH, and 0.6 mg of glycine, was dissolved in 1 mL of 0.1 M Tris-HCl, pH 7.2, containing 2.5% SDS and 10% mercaptoethanol, boiled 5 min, cooled to 25 °C, and applied to the column (1.5 × 86.5 cm). Flow rate was 18 mL/h, and 2-mL fractions were collected and assayed for absorbance at 280 nm, radioactivity (cpm), SH groups (Novoa *et al.*, 1966), and ninhydrin reactivity (Hirs, 1967). The mass of each of the radioactive peptides was estimated by comparison with marker proteins and peptides (Figure 1A). The distribution coefficient (K_d) was calculated from the equation, $K_d = (V_e - V_o)/(V_i - V_o)$, taking V_o, V_i, and V_e as the volumes of elution of BSA, mercaptoethanol, and the compound of interest, respectively.

In a second method (Carnegie, 1965), the samples, mixed with 1.5 mg cytochrome *C*, 1 mg of glucagon, 1 mg of bradykinin, and 0.4 mg of alanine were lyophilized, dissolved in 0.6 mL of phenol−acetic acid−water (1:1:1, w/v/v), and applied to a Sephadex G-50 (fine) column (1 × 50 cm), which was equilibrated and eluted (flow rate, about 10 mL/h) by gravity with the same solvent, collecting fractions of about 1 mL. The weight, the radioactivity, and the ninhydrin reactivity (Hirs, 1967) of the fractions were determined. Ninhydrin assay was done before and after alkaline hydrolysis of the samples. For alkaline hydrolysis, the solvent was evaporated (6 h at 120 °C) and the residue was dissolved in 0.15 mL of 13.5 N NaOH and incubated 1 h at 100 °C. Samples were adjusted to pH 5 with either acetic acid (after alkaline hydrolysis) or NaOH (no hydrolysis) before ninhydrin assay. For the estimation of K_d values (Figure 1B), V_o and V_i were taken as the volumes of elution of cytochrome *C* and alanine, respectively.

Cleavage of Asn-Gly Bonds with Hydroxylamine. 0.01 mg of the 25.8-kDa COOH-terminal fragment was incubated 4 h at 45 °C in 0.1 mL of hydroxylamine solution (Bornstein & Balian, 1977). The digestion was terminated with 14 μL formic acid and 0.4 mL of 60% acetonitrile containing 0.1% trifluoracetic acid. The mixture was applied to a Sephadex G-25 column (1.5 × 5 cm) which was equilibrated and eluted with the acetonitrile solution. The radioactive material was recovered in the excluded volume and was lyophilized, dissolved with heating at 100 °C in a solution of 10 mM sodium phosphate, pH 7.2 containing 7 M urea, 1% SDS, 1% 2-mercaptoethanol and 0.01% bromophenol blue, and subjected to SDS−urea−PAGE using lower and upper gels of 15% and 3.5% polyacrylamide, respectively, containing 0.1 M sodium phosphate, pH 7.2, 0.1% SDS, and 6 M urea [Shapiro *et al.*, (1967), as modified in BRL Catalogue 1981/82]. These gels are suitable for separation of low molecular weight proteins. The gels were fixed in a solution of 10% acetic acid, 10% trichloroacetic acid, 30% methanol, 0.1% $CuSO_4$, and 6% formaldehyde and were used for fluorography (Bonner & Laskey, 1974). Alternatively, the unfixed gels were cut perpendicularly to the migration axis in 3.5-mm slices and solubilized by 12-h incubation at 60 °C with 0.6 mL of a solution of 20% $HClO_4$ and 20% H_2O_2, and the radioactivity in each slice was determined.

CNBr Cleavage. The 25.8-kDa COOH-terminal fragment (0.01 mg/mL) was incubated with 0.24 mM CNBr in 70% formic acid at 24 °C (Charbonneau, 1989). After the indicated periods, 0.1-mL samples were mixed with 5 μL of a solution of 0.1 mg of BSA/mL, lyophilized, and the residue was subjected to SDS−urea−PAGE.

For NH_2-terminal sequencing of the 13-kDa CNBr product, the 25.8-kDa fragment (0.25 mg) was digested with 600 μmol of CNBr in 0.35 mL of 70% formic acid for 13 h, and the reaction was terminated by injecting the entire sample to the HPLC system.

Extensive Digestions with Proteases. Except where indicated, the incubations were done at 25 °C for 24 h in a volume ≤0.5 mL, the amount of peptide substrate was ≤50 μg, and the protease was added at the start and again after

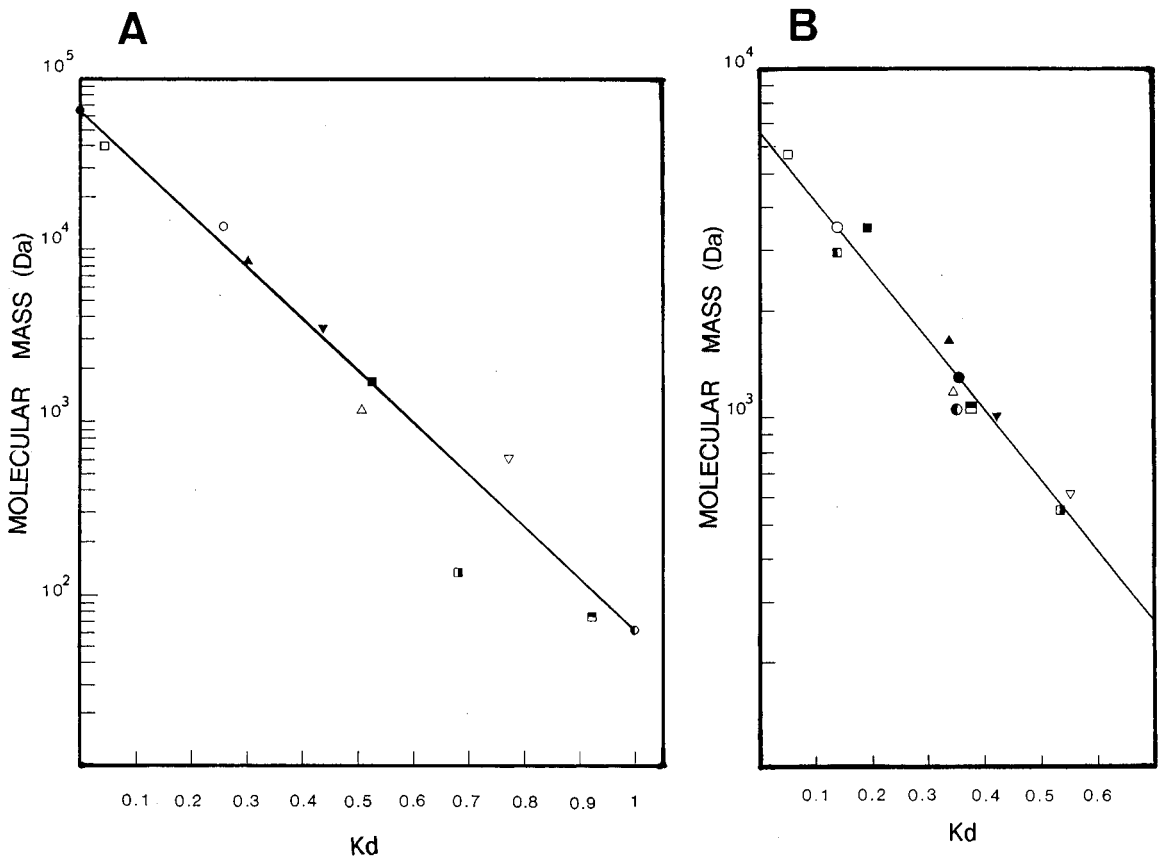

FIGURE 1: Plots used for the estimation of the molecular weight of peptides by gel filtration on (A) Sephacryl S-200 in the presence of SDS and (B) Sephadex G-50 in phenol:acetic acid:water (1:1:1). (A) ●, BSA; □, aldolase; ○, ribonuclease A; ▲, ubiquitin; ▼, glucagon; ■, somatostatin; △, GnRH; ▽, oxidized glutathione; ◨, glycyl-glycine; ⬒, glycine; ◐, 2-mercaptoethanol. (B) □, insulin; ■, chain B of insulin; ○, glucagon; ◧, ACTH; ▲, somatostatin; △, angiotensin I; ⬒, GnRH; ◐, angiotensin II; ●, bradykinin; ▼, oxytocin; ▽, leu-enkephalin; ◨, oxidized glutathione.

12 h. Digestions were terminated by injection into the HPLC system. The radioactive material remained unchanged (shown by HPLC) in control incubations without protease. The 25.8-kDa fragment was digested with trypsin (20 μg/mL per addition) at 37 °C in 0.1 M NH_4HCO_3 containing 2 M urea. Digestions with V8 protease (10 μg/mL, with up to five readditions at 12-h intervals) were done in 0.21 M Tris-HCl, pH 7.4, in the presence of 10% acetonitrile and/or 0.1% SDS. Alternatively, 50 mM NH_4HCO_3, pH 7.8, containing 2 M urea was used. Tryptic digestion of the minimal V8 protease ^{14}C-labeled fragment was carried out at 37 °C using 1 μg of trypsin/mL, with readditions after 5 and 12 h. Digestions of the 25.8-kDa fragment with endoproteinase Arg-C (0.2 mg/mL) or Lys-C (7.5 μg/mL) were carried out for 20 h in a solution of either 0.22 M Tris-HCl, pH 7.4, containing 2 mM dithioerythritol, 1 mM $CaCl_2$, and 10% acetonitrile, or, for endoproteinase Lys-C, in 0.25 M Tris-HCl, pH 8.6. The same results were obtained using 20 μg of endoproteinase Lys-C/mL at 37 °C, in a solution of 25 mM Tris-HCl, pH 8.5, containing 1 mM EDTA and 20 mM methylamine.

Isolation of [^{14}C]Peptides for NH_2-Terminal Sequencing. In order to selectively digest the COOH-terminal 18 kDa of the large subunit from the enzyme molecules that do not incorporate UMP (about 95% of the enzyme; Rubio et al., 1991), the enzyme (1.4 mg/mL), after photoaffinity labeling, was incubated with trypsin (1.25 μg/mL; bovine, from Boehringer Mannheim), at 37 °C, in 0.12 M sodium phosphate, pH 7.7. The digestion was stopped after 12 min by adding 2 mM PMSF and boiling for 5 min in 0.5% SDS. The protein was then placed in 0.125 M Tris-HCl, pH 6.8, containing 0.5% SDS, by passage through Sephadex G-25 (PD-10 column; from Pharmacia). Then, the COOH-terminal fragment of 25.8 kDa was generated exclusively from the UMP-labeled enzyme molecules (the enzyme molecules that remain uncleaved after the limited tryptic digestion) by incubation with V8 protease (10 μg/mL) for 25 min at 37 °C. The incubation was terminated by heating 5 min at 100 °C, and the 25.8-kDa fragment was isolated (Rubio *et al.*, 1991). About 1 nmol of this fragment was digested with trypsin (1.2 μg) at 35 °C for 24 h, in 0.2 mL of 0.1 M NH_4HCO_3 containing 20 mM methylamine and 2 M urea (Stone *et al.*, 1989). The resulting "minimal" [^{14}C]-peptide was isolated by HPLC using a 60-min linear gradient (9.5%−38%) of acetonitrile in 0.1% trifluoroacetic acid.

During HPLC isolation of the 25.8-kDa fragment prepared selectively from the [^{14}C]UMP-labeled enzyme molecules (see previous paragraph), about 15% of the radioactivity was eluted as an early peak (30% acetonitrile). This fraction consisted of the "minimal" [^{14}C]tryptic peptide (see Results). About 0.5 nmol of this peptide was digested with V8 protease (1 μg) for 24 h at 25 °C in 0.5 mL of 40 mM NH_4HCO_3, pH 7.8, and the resulting labeled peptide was isolated by HPLC, using a 60-min linear gradient (14%−52%) of acetonitrile in 0.1% trifluoroacetic acid.

The amino-terminal sequences were determined at the Servicio de Secuenciación of the University of Barcelona by automated Edman degradation with an Applied Biosystems 470A Protein Sequencer. The Pth derivatives were

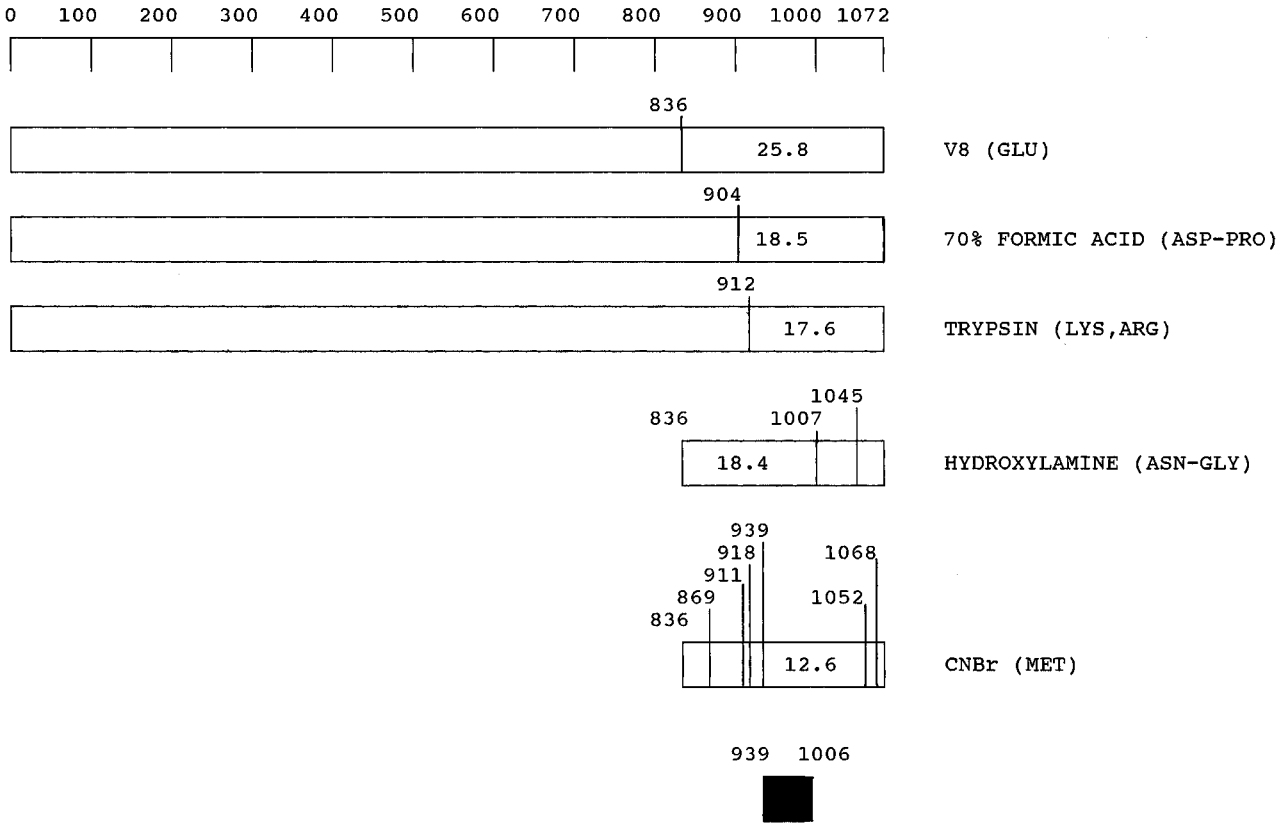

FIGURE 2: Strategy used to identify the region where UMP is bound. The scale corresponds to the numbering of residues in the large subunit of *E. coli* carbamyl phosphate synthetase (Nyunoya & Lusty, 1983). The horizontal bars represent the large subunit or its fragments, indicating the number of the NH_2-terminal residue of each fragment and the mass in kDa (as deduced from the sequence) of the fragments where UMP has been shown to be attached. The cleavage techniques used are indicated to the right, giving in parentheses the specificity of the cleavage. The black box at the bottom represents the minimal region within which UMP is attached. The cleavage sites of V8 staphylococcal protease or 70% formic acid were reported by Rubio *et al.* (1991). The site of cleavage by trypsin was identified by Bueso *et al.* (1994).

analyzed on an Applied Biosystems 120A on-line HPLC system using a microbore C-18 Brownlee column.

RESULTS

Mapping of the UMP Site by Chemical Cleavage. In earlier photolabeling studies we showed (Rubio *et al.*, 1991; Bueso *et al.*, 1994) that UMP is attached to the COOH-terminal domain downstream of Arg911 (Figure 2). In the present experiments, we use hydroxylamine or CNBr to localize in more detail the site of attachment of [^{14}C]UMP in the COOH-terminal 25.8-kDa fragment generated by limited digestion of the enzyme with V8 protease. This fragment (residues 836−1072) contains two Asn-Gly bonds at residues 1006−1007 and 1044−45 (Nyunoya & Lusty, 1983). Hydroxylamine, by cleaving at these bonds, should generate products of 18.4, 4.4, and 2.9 kDa (Figure 2). Figure 3 shows that after cleavage with hydroxylamine, the radioactivity migrates in SDS−urea−PAGE according to a mass of 18.4 kDa, indicating that UMP is attached to the large fragment. No labeled fragment smaller than 18.4 kDa was detected by fluorography or by scintillation counting of a sliced unfixed gel. The presence of a faint band of 23 kDa is consistent with partial cleavage at Asn1044 only.

The 25.8-kDa fragment contains 6 methionines (Figure 2), so that CNBr digestion should produce one 12.6-kDa fragment and six smaller (<5 kDa) fragments. After treatment with CNBr 90% of the radioactivity migrated in SDS−urea−PAGE with a mass of 12 kDa (Figure 3),

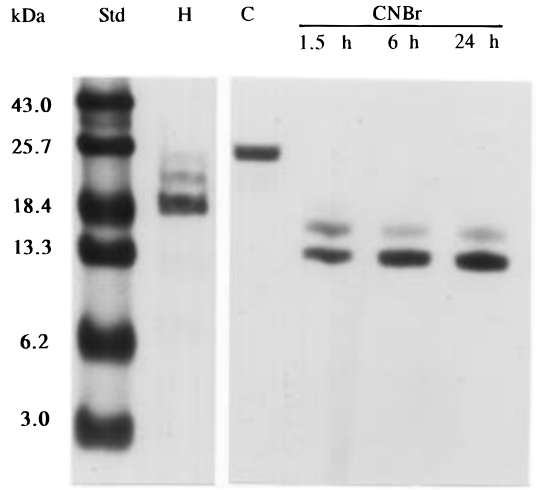

FIGURE 3: Cleavage of the UMP-labeled 25.8-kDa fragment by hydroxylamine (H) or CNBr, analyzed by SDS−urea−PAGE and fluorography. The figures at the top denote the number of hours of incubation with CNBr. The period of incubation with hydroxylamine was 4 h. C, undigested control. Std, ^{14}C-labeled molecular weight standards. The masses of the standards are shown to the left. For further details see Experimental Procedures.

indicating that UMP is attached to the 12.6-kDa fragment. NH_2-terminal sequencing (see Experimental Procedures) confirmed that cleavage occurred at Met938. The faint band of 15 kDa observed (Figure 3, CNBr), particularly with short incubations, reveals some partial cleavage either at Met938 or at Met1051. Taken together, the results of cleavage with

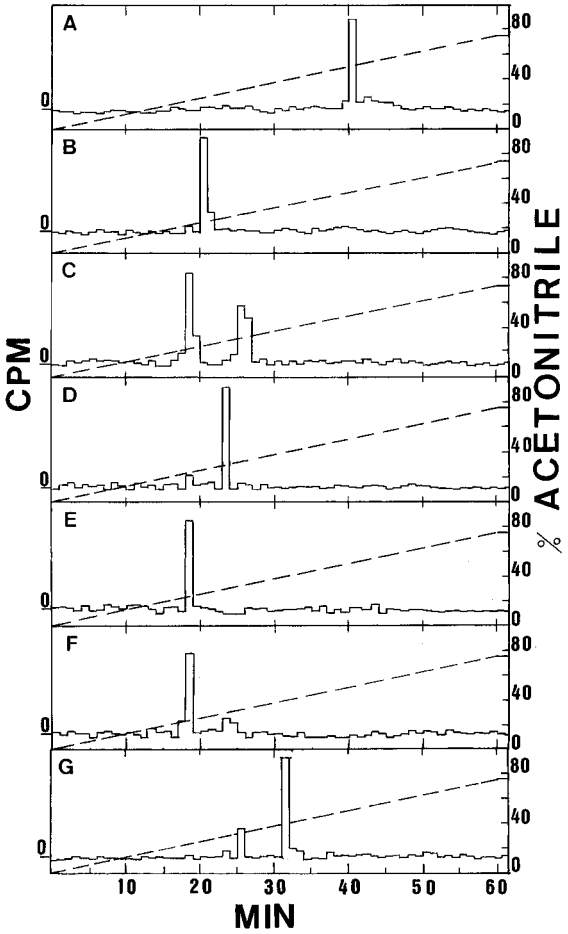

FIGURE 4: HPLC elution profiles of the radioactive products of exhaustive digestion with different proteases of the 25.8-kDa COOH-terminal fragment generated from the UMP photoaffinity labeled enzyme. (A) 25.8-kDa fragment prior to digestion and (B, D, G) after digestion with trypsin, V8 staphylococcal protease, or endoproteinase Lys-C, respectively. (C) Digestion with V8 protease in the presence of 10% acetonitrile and 0.1% SDS of the material eluted at min 21 in B. (E, F) Digestion with trypsin and with endoproteinase Arg-C of the material eluted at min 24 in D. The broken line represents the concentration of acetonitrile. The actual number of counts is not shown for simplicity and for superimposition of replicate experiments (at least two per profile) in which up to 100-fold different amounts of radioactivity (range, 100−10 000 cpm in the highest peak) were used.

hydroxylamine and with CNBr show that UMP is attached to the protein between residues 939 and 1006 (Figure 2).

Peptide Mapping of the Site. When the 25.8-kDa fragment was digested with trypsin, a single radioactive product was obtained. The product eluted early from the reverse-phase column (Figure 4B), and exhibited a mass of about 2.5 kDa (Table 1). The 2.5-kDa labeled peptide was susceptible to cleavage by V8 protease (Figure 4C), and when SDS and acetonitrile were present yielded radioactive products of about 0.5 and 2.25 kDa (Table 1). However, under conditions which ensured cleavage solely at glutamate residues (Drapeau, 1977), only the 0.5-kDa product was released. Digestion of the 25.8-kDa fragment with V8 protease yielded a single radioactive fragment (Figure 4D) of about 1.5 kDa (Table 1), which was susceptible to cleavage both by trypsin and by endoproteinase Arg-C (Figure 4E,F), yielding the 0.5-kDa peptide with either of the two proteases (Table 1). These results indicate that the 2.5-kDa tryptic peptide and the 1.5-kDa V8 fragment contain internal glutamate and arginine, respectively. Further, the 0.5-kDa peptide corresponds to the region of overlap shared by the two longer fragments.

Table 1: HPLC Retention Time and Mass Estimate by Two Gel Filtration Methods of the Radioactive Peptides Generated by Exhaustive Proteolytic Digestion of the 25.8-kDa [^{14}C]UMP-Labeled Fragment

protease (s)	retention time (min)	molecular mass (Da) S-200	G-50	$\bar{X}$
none	41	25500	>6600	
trypsin	21	2731	2193	2462
trypsin + V8	19	355	650	503
trypsin + V8 + SDS	26.5	2566	1936	2251
V8	24	2085	1005	1545
V8 + trypsin	19	472	596	534
V8 + endoproteinase Arg-C	19	417	647	532
endoproteinase Lys-C	32	5231	4874	5052
(minor peak)	26	3551	3213	3382
endoproteinase Lys-C[a] + V8	24	2042	1089	1566

[a] V8 digestion of the major product (retention time, 32 min) of endoproteinase Lys-C digestion.

Endoproteinase Lys-C is expected to generate few fragments (Table 2) from the enzyme region (residues 939−1006; Figure 2) where UMP is attached, and thus the identification of the resultant labeled products is more straightforward than with other proteases that cleave more frequently. Digestion of the 25.8-kDa fragment with endoproteinase Lys-C yielded a major late-eluting radioactive product having a mass of about 5 kDa and an earlier minor (<30% of the radioactivity) component of approximately 3.4 kDa (Figure 4G and Table 1). From the large mass of these fragments we confirmed that trypsin cleaves a bond involving arginine to generate the 2.5-kDa fragment. The 3.4-kDa fragment generated with endoproteinase Lys-C might correspond to either peptide K−V or K−VII (Table 2). Attachment of UMP to K−VII is considered unlikely, because Asn1006 is part of the Asn-Gly bond which is cleaved by hydroxylamine and is the only residue of K−VII that is located within the UMP-labeled region (residues 939−1006). Furthermore, photochemical addition of UMP to Asn1006 would prevent hydroxylamine cleavage of this Asn-Gly bond, which is clearly not the case. The endoproteinase Lys-C 5-kDa product may correspond to K−V + K−VI (Table 2), which would be generated if the proteinase was unable to cleave the bond at Lys992. As predicted from the sequence of K−V + K−VI, digestion of the 5-kDa fragment with V8 protease yielded the identical 1.5-kDa labeled peptide observed with V8 digestion of the 25.8-kDa fragment (data not shown). Only the minimal V8 peptide (983)AGINPRLVNKVHE meets all the requirements demanded of K−V + K−VI by the experimental results. A part of the sequence (underlined) is found in K−V; its calculated mass is close to the observed value of 1.5 kDa, it contains internal arginine, and its tryptic products include a peptide of about 0.5 kDa, (983)AGINPR, whose sequence is preceded by glutamate and terminated with arginine. Moreover, if the bond involving Lys992 is resistant to tryptic cleavage, then the tryptic digestion would also give the peptide (989)LVNKVHE, which also has a mass close to 0.5 kDa and is bounded by arginine and by glutamate. Sequencing studies (see next section) confirmed the resistance to trypsin of the bond involving Lys992.

Table 2: Predicted Minimal Endoproteinase Lys-C Peptides That Constitute the UMP-Labeled Region (Residues 939−1006)[a]

peptide	N-terminal residue	amino acid sequence	molecular mass[b] (Da)
K−I	930	AQLGSNSTM*KK	1036
K−II	941	HGRALLSVREGDK	1419
K−III	954	ERVVDLAAK	1000
K−IV	963	LLK	355
K−V	966	QGFELDATHGTAIVLGEAGINPRLVNK	2822
K−VI	993	VHEGRPHIQDRIK	1585
K−VII	1006	N*GEYTYIINTTSGRRAIEDSRVIRRSALQYK	3459

[a] The peptides that would result from cleavage at lysine of the sequence of the *E. coli* enzyme (Nyunoya & Lusty, 1983) are shown, ordered from more amino- to more carboxyl-terminal. The peptides cover the entire region delimited between the sites of CNBr and hydroxylamine cleavage (indicated by asterisks) that set the boundaries to the UMP-labeled region. [b] Molecular masses calculated from the published sequence.

NH_2-Terminal Sequence of UMP-Labeled Peptides. Since only ~5% of the enzyme molecules incorporate UMP upon UV-irradiation (Rubio *et al.*, 1991), we found that it is essential to separate the UMP-labeled and unlabeled peptides before amino acid sequence analysis. By exploiting the differences between UMP-labeled and unlabeled enzyme molecules in their susceptibility to limited tryptic digestion (Rubio *et al.*, 1991), we isolated after limited sequential digestions with trypsin and V8 protease (see Experimental Procedures), the 25.8-kDa COOH-terminal fragment with about 20-fold increased specific radioactivity (cpm:A_{214} ratio). This purified UMP-labeled fragment was digested with trypsin, yielding the UMP-labeled 2.5-kDa peptide that was used for NH_2-terminal sequencing. The sequences STGEV-MGVGR and LVNXVHEGRPHIQD (X indicates that no Pth derivative was observed) were obtained and were identified in the primary structure of the large subunit, beginning, respectively, at residues 912 and 989, and following in both cases an arginine residue. Only the peptide (989)LVNKVHEGRPHIQD is within the region (residues 939−1006) shown to be labeled with UMP (Figure 2) and contains the sequence (989)LVNK, which we found (see previous section) to be one of the two candidate sites for UMP attachment. The failure to identify Lys992 in this peptide by sequence analysis, the lack of cleavage by trypsin at this lysine, and the partial cleavage by endoproteinase Lys-C (see previous section) indicate that Lys992 is covalently attached to UMP. The bond in Arg-Pro, which was also uncleaved in the 2.5-kDa peptide, is known to be resistant to trypsin (Allen, 1989).

As an unexpected secondary product of the HPLC separation of the UMP-labeled 25.8-kDa fragment, we also isolated an early-eluting radioactive peptide (see Experimental Procedures), which, upon characterization (HPLC analysis, mass, and susceptibility to proteases), was found to be the 2.5-kDa tryptic fragment, highly enriched in the UMP-labeled form. Thus, in contrast to the majority of the UMP-labeled enzyme molecules, some (perhaps, those photolytically damaged) of the labeled molecules appear to be highly susceptible to trypsin, yielding, already in the initial tryptic treatment, the UMP-labeled 2.5-kDa fragment. We digested the 2.5-kDa fragment with V8 protease, and the resulting UMP-labeled 0.5-kDa peptide was isolated and sequenced. The NH_2-terminal sequence obtained, LVNXVHE, is, as expected, a part of the sequence of the 2.5-kDa tryptic peptide. Even though the Pth derivatives in cycles 5−7 gave a very good signal in the HPLC analysis, no other amino acid derivatives were detected in three additional cycles of sequencing. As in sequence analysis of the 2.5-kDa tryptic peptide, Lys992 was not detected. Consistent with the absence of Pth-Lys in cycle 4, the radioactivity was retained in the membrane.[2] These results confirm that UMP is attached to the peptide (989)LVNK as a covalent adduct of lysine 992.

DISCUSSION

Our present results show that photochemical reaction of carbamyl phosphate synthetase with the inhibitor UMP results in the formation of a covalent adduct of the nucleotide and a single amino acid residue (lysine 992) located in the COOH-terminal domain of the large subunit. In model studies (Schott & Shetlar, 1974) lysine is found to be highly reactive in the photochemical addition to uridine, and detailed studies of the photochemical reactions of 1,3-dimethyluracil with 1-aminopropane (used as an analog of the lysine side chain) demonstrated the formation of four products resulting from addition of either the amino nitrogen or carbon 1 of the amine to carbons 5 and 6 of the uracil ring (Gorelic *et al.*, 1972). More than one product might also be formed in the photochemical addition to UMP of lysine 992 of carbamyl phosphate synthetase, because even though the modified lysine becomes entirely refractory to cleavage by trypsin, a small but substantial fraction is still cleaved by endoproteinase Lys-C, indicating that at least one adduct is still recognized as lysine by this protease.

The large subunit of *E. coli* carbamyl phosphate synthetase (Nyunoya & Lusty, 1983) contains 56 lysines and at least 163 additional residues capable of photochemical addition to uracil (Schott & Shetlar, 1974). The fact that only lysine 992 adds to UMP demonstrates the extreme specificity of the reaction, implying that lysine 992 is located in the binding site for UMP and that it interacts with the pyrimidine ring. The absence of lysine in the equivalent position of hamster CAD (Figure 5, marked with an arrow) not only explains the poor photoaffinity labeling of this protein with its inhibitor UTP (V. Rubio and E. A. Carrey, unpublished experiments) but also shows that lysine itself is not essential for the binding of the pyrimidine inhibitor. Possibly, the lysine residue contributes to the affinity of binding of the nucleotide, because the K_i for the pyrimidine inhibitor (Anderson & Meister, 1966; Paulus & Switzer, 1979) is significantly lower in the *E. coli* and the pyrimidine-specific *Bacillus subtilis* carbamyl phosphate synthetases that contain this lysine (Figure 5), than in the pyrimidine-specific enzymes from hamster (Liu *et al.*, 1994) and yeast (Denis-Duphil, 1989), which lack the analogous lysine.

[2] The amount of the Pth-Lys992-UMP adduct was too small for further chemical or mass spectrometry characterization.

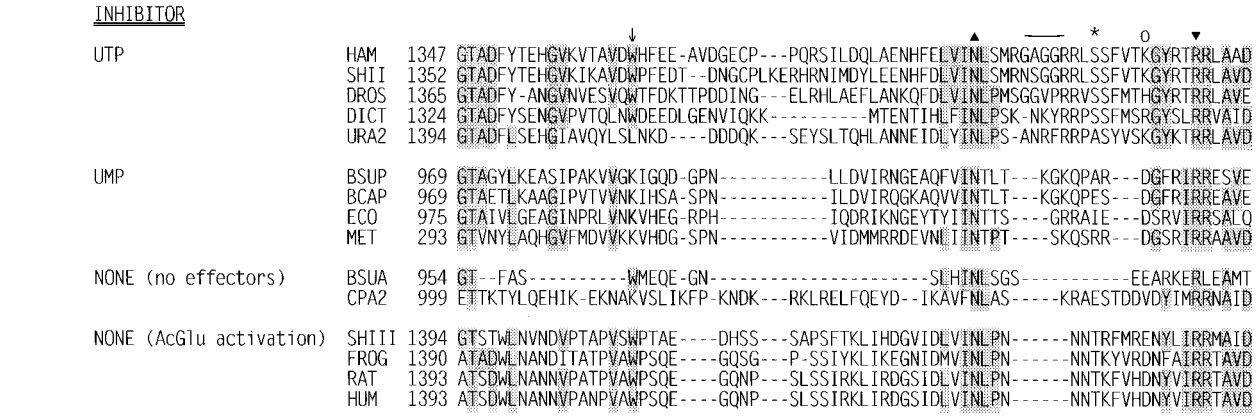

FIGURE 5: Alignment of carbamyl phosphate synthetase sequences in the region of the putative effector binding fold. The alignment is based on and extends that of Simmer *et al.* (1990). Shadowing indicates conservation of a residue in at least eight of the 15 sequences. The arrow signals the residue corresponding in the different carbamyl phosphate synthetases to lysine 992 of *E. coli*. The symbols ▲ and ▼ denote an invariant asparagine and an invariant arginine, respectively. The horizontal line and the asterisk indicate, respectively, the glycine-rich sequence GAGGR and the serine that is phosphorylated by cAMP-dependent protein kinase (Carrey *et al.* 1985) in hamster CAD. The small circle indicates a basic residue found in the UTP-inhibited enzymes that is replaced by an acidic residue in the UMP-inhibited or acetylglutamate-activated enzymes. Unless indicated, the sequences were taken from the Swiss-Prot data bank, release 31. HAM hamster CAD; SHII, CAD from shark (Hong *et al.*, 1995); DROS, *D. melanogaster CAD* [in this case the conserved residues 1365−1367 are given according to Simmer *et al.* (1990), who used for these residues a reading frame differing from that used in the original publication (Freund & Jarry, 1987)]; DICT, the pyrimidine-specific enzyme from *Dictyostelium discoideum* (Faure *et al.* 1989); URA2, the product of the URA2 gene from *Saccharomyces cerevisiae*; BSUP and BCAP, the pyrimidine-specific enzymes from *B. subtilis* and *Bacillus caldolythicus*, respectively; ECO, carbamyl phosphate synthetase from *E. coli*; MET, carbamyl phosphate synthetase from *M. barkeri* (in this case only part of the sequence of the enzyme is available); BSUA and CPA2, arginine-specific carbamyl phosphate synthetases from *B. subtilis* and *S. cerevisiae*, respectively; SHIII, carbamyl phosphate synthetase III from shark (Hong *et al.*, 1994); FROG, RAT and HUM, carbamyl phosphate synthetase I from frog (Helbing & Atkinson, 1994), rat, and human, respectively.

On the basis of the alignment of the sequences (Figure 5) surrounding the residue corresponding to the labeled lysine (marked with an arrow in Figure 5), the carbamyl phosphate synthetases fall into distinct groups, which reflect the nature of their allosteric effectors. In the pyrimidine-specific enzymes from *Bacilli* and in the synthetase from *E. coli*, which are inhibited by UMP, the lysine is present in the sequence (V/L)VXK(I/V). This sequence is also present in *Methanosarcina barkeri* carbamyl phosphate synthetase, suggesting that this uncharacterized bacterial enzyme is involved in pyrimidine biosynthesis and is inhibited by UMP. In contrast, in the arginine-specific *B. subtilis* enzyme, an enzyme that is not inhibited by UMP, the sequence (V/L)-VXK(I/V) is missing. In this enzyme the lysine is replaced by tryptophan, which is preceded by a 10-residue deletion. In addition, there is an even longer deletion beginning eight residues downstream of the tryptophan and ending four residues upstream of an invariant asparagine residue found in all carbamyl phosphate synthetases of known sequence (marked with the symbol ▲ in Figure 5). Clearly, the deleted regions are not essential for catalysis or for the structural stability of the protein. Since the arginine-specific *B. subtilis* enzyme is not subject to regulation by allosteric effectors, the deleted regions represent loss of the binding site for the nucleotide effector.

As in the arginine-specific enzyme from *Bacilli*, the arginine-specific enzyme from yeast is also not modulated by allosteric effectors (Price *et al.*, 1978). In this case the alignment (CPA2, Figure 5) shows that while the lysine is conserved, the (V/L)V residues that are found upstream of lysine in the signature sequence of the UMP-regulated enzymes, are not conserved in the yeast enzyme, suggesting that (V/L)V is important for binding of the effector.

In the carbamyl phosphate synthetases from multicellular organisms (including *Dictyostelium discoideum*) the lysine is replaced by a tryptophan (Figure 5). In the pyrimidine-specific enzymes from these organisms, the tryptophan is located within the context (V/L)XWXX(D/E) and the pyrimidine inhibitor is UTP rather than UMP. In the pyrimidine-specific yeast enzyme (Ura2) the corresponding sequence agrees with the signature sequence, except for the replacement of lysine by leucine rather than tryptophan. It appears that the use of UTP rather than UMP is associated with the replacement of the lysine by a residue of increased hydrophobicity. In the ureotelic or ureosmotic enzymes, which are insensitive to nucleotide effectors but which are activated by acetylglutamate, the tryptophan found in multicellular organisms is part of another signature sequence, PV(A/S)WP(T/S)(A/Q)E. The conservation of this signature from shark to man suggests that this sequence is functionally important in this group of enzymes, perhaps for the binding of acetylglutamate. In fact, this signature sequence is within the region of the rat liver enzyme (residues 1351−1467) that reacts with the acetylglutamate analog, *N*-chloroacetyl-L-glutamate, in photoaffinity labeling experiments (Bendala, 1993).

A second region of the C-terminal domain that appears to be involved with binding of nucleotide effectors was identified in CAD by the experiments of Carrey *et al.* (1985), who showed that phosphorylation of serine 1406 (marked with an asterisk in Figure 5) reduces inhibition by UTP. The sequence of the *E. coli* enzyme corresponding to the phosphorylation site in CAD is located 30 residues downstream of lysine 992, centered between the invariant asparagine and an invariant arginine that is found in all carbamyl phosphate synthetases sequenced thus far (marked with the symbol ▼ in Figure 5). Both the invariant asparagine and arginine are flanked by highly conserved residues. The sequence between the two conserved clusters is five to seven residues longer in the carbamyl phosphate synthetases that

are inhibited by UTP than in those that are inhibited by UMP or that are insensitive to nucleotides. It is reasonable to expect this region to be longer in the enzymes that are inhibited by UTP than in those that are inhibited by UMP, if the region accommodates the polyphosphate chain of UTP, as suggested by the phosphorylation results with CAD. Loops that accommodate the polyphosphate chain of polynucleotides are flexible and generally are enriched with glycine (Yamaguchi *et al.*, 1993). Hamster CAD contains the sequence GAGGR (marked in Figure 5 with a horizontal line above the sequence) with these characteristics. This glycine-rich sequence is located between the invariant asparagine and the phosphorylation site. Similar sequences were also disclosed in the enzymes from shark (type II) and *Drosophila melanogaster* (Figure 5). In the case of yeast and *D. discoideum*, even though no glycine is present, the proposed loop consists of four small or polar residues with high β-turn potential. In the group of enzymes that bind UTP the sequence between the invariant asparagine and arginine is rich in basic residues and contains no acidic residues, consistent with an involvement in the neutralization of the negative charges of the polyphosphate chain of UTP. Of possible interest is the basic residue K/H/R (indicated with an "0" in Figure 5) which is only found among the enzymes that are inhibited by UTP.

The bacterial carbamyl phosphate synthetases that are inhibited by UMP contain neither a sequence equivalent to the glycine-rich sequence nor a phosphorylation site. The sequences contain at least one acidic residue and are generally poorer in basic residues than the UTP-inhibited enzymes. This is also true of the arginine-specific enzymes. It is significant that both the proposed polyphosphate binding loop and the phosphorylation site, and in fact most of the region between the invariant asparagine and arginine has been deleted in the arginine-specific *B. subtilis* enzyme, an enzyme that has no known effectors.

In summary, our identification of lysine 992 as the site of photochemical addition of UMP in *E. coli* carbamyl phosphate synthetase, and the analysis of the sequences surrounding this lysine provide strong evidence for the location of the binding fold for the nucleotide inhibitor. The fold comprises two contiguous regions. The first region surrounds lysine 992, extending downstream to the cluster of conserved residues containing an invariant asparagine. This region presumably interacts with the pyrimidine ring of the inhibitor. The second region extends from the invariant asparagine to an invariant arginine residue and possibly interacts with the phosphate moiety of the nucleotide, at least in the enzymes that are inhibited by UTP. Multiple sequence alignment of carbamyl phosphate synthetases (Simmer *et al.*, 1990; extended to all the enzymes of Figure 5) shows that the two regions of the putative uridine nucleotide binding fold are those regions of the COOH-terminal domain that exhibit the highest sequence variability amongst the different carbamyl phosphate synthetases. This suggests that modifications in the primary structure of this fold provide a basis for the different regulatory properties of the various carbamyl phosphate synthetases.

After submission of this manuscript, a paper appeared (Czerwinski *et al.*, 1995) demonstrating that the effects of UMP are lost when the large subunit is truncated at residue 981. This finding is consistent with our data, since both lysine 992 and the entire UMP binding fold proposed here are located within the deleted region. Czerwinski *et al.* (1995) also found that enzyme forms with mutations at threonine 976 or at glycine 967 become insensitive to UMP, this effect being highly selective for UMP in the T976A mutant. Although the binding of UMP was not assayed in these studies, the fact that these residues are only 16 and 25 residues from lysine 992 agrees with our localization of the UMP binding fold in this region of the protein.

ACKNOWLEDGMENT

We thank Drs. M. A. Miranda (Universidad Politecnica de Valencia) and M. D. Shetlar (University of California) for advice on photochemistry and C. Buesa, J. A. Ortiz and the Servicio de Secuenciación de la Universidad de Barcelona for determining the sequences. V. Rubio and J. Cervera are members of the IIC-KUMC International Molecular Cytology Programme.

REFERENCES

Allen, G. (1989) *Sequencing of Proteins and Peptides*, 2nd ed., Elsevier, Amsterdam.
Anderson, P. M., & Meister, A. (1966) *Biochemistry 5*, 3164−3169.
Belew, M., Fohlman, J., & Janson, J. C. (1978) *FEBS Lett. 91*, 302−304.
Bendala, E. (1993) Ph.D. Thesis, University of Valencia, Valencia, Spain.
Bonner, W. M., & Laskey, R. A. (1974) *Eur. J. Biochem. 46*, 83−88.
Bornstein, P., & Balian, G. (1977) *Methods Enzymol. 47*, 132−145.
Bueso, J., Lusty, C. J., & Rubio, V.(1994) *Biochem. Biophys. Res. Commun. 203*, 1083−1089.
Carnegie, P. R. (1965) *Nature 206*, 1128−1130.
Carrey, E. A., Campbell, D. G., & Hardie, D. G. (1985) *EMBO J. 4*, 3735−3742.
Cervera, J., Conejero-Lara, F., Ruiz-Sanz, J., Galisteo, M. L., Mateo, P. L., Lusty, C. J., & Rubio, V. (1993) *J. Biol. Chem. 268*, 12504−12511.
Charbonneau, H. (1989) in *A Practical Guide to Protein and Peptide Purification for Microsequencing* (Matsudaira, P. T., Ed.) pp 15−30, Academic Press, New York.
Czerwinski, R. M., Mareya, S. M., & Raushel, F. M. (1995) *Biochemistry 34*, 13920−13927.
Denis-Duphil, M. (1989) *Biochem. Cell Biol. 67*, 612−631.
Drapeau, G. R. (1977) *Methods Enzymol. 47*, 189−191.
Evans, D. R., & Balon, M. A. (1988) *Biochim. Biophys. Acta 953*, 185−196.
Faure, M., Camonis, J. H., & Jacquet, M. (1989) *Eur. J. Biochem. 179*, 345−358.
Freund, J. N., & Jarry, B. P. (1987) *J. Mol. Biol. 193*, 1−13.
Gorelic, L. S., Lisagor, P., & Yang, N. C. (1972) *Photochem. Photobiol. 16*, 465−480.
Guadalajara, A. M. (1987) Ph.D. Thesis, University of Valencia, Valencia, Spain.
Hall, L. M., Metzenberg, R. L., & Cohen, P. P. (1958) *J. Biol. Chem. 230*, 1013−1021.
Helbing, C. C., & Atkinson, B. G. (1994) *J. Biol. Chem. 269*, 11743−11750.
Hirs, C. H. W. (1967) *Methods Enzymol. 11*, 325−329.
Hong, J., Salo, W. L., Lusty, C. J., & Anderson, P. M. (1994) *J. Mol. Biol. 243*, 131−140.
Hong, J., Salo, W. L., & Anderson, P. M. (1995) *J. Biol. Chem. 270*, 14130−14139.
Liu, X., Guy, H. I., & Evans, D. R. (1994) *J. Biol. Chem. 269*, 27747−27755.
Marina, A., Bravo, J., Fita, I., & Rubio, V. (1995) *Proteins: Struct., Funct., Genet. 22*, 193−196.
Marshall, M., & Fahien, L. A. (1988) *Arch. Biochem. Biophys. 262*, 455−470.

Meister, A. (1989) *Adv. Enzymol. Relat. Areas Mol. Biol. 62*, 315−374.
Novoa, W. B., Tigier, H. A., & Grisolia, S. (1966) *Biochim. Biophys. Acta 113*, 84−94.
Nyunoya, H., & Lusty, C. J. (1983) *Proc. Nat. Acad. Sci. U.S.A. 80*, 4629−4633.
Paulus, T. J., & Switzer, R. L. (1979) *J. Bacteriol. 137*, 82−91.
Powers-Lee, S. G., & Corina, K. (1986) *J. Biol. Chem. 261*, 15349−15352.
Price, C. W., Holwell, J. H., & Abdelal, A. T. H (1978) *J. Gen. Microbiol. 106*, 145−151.
Rodríguez-Aparicio, L. B., Guadalajara, A. M., & Rubio, V. (1989) *Biochemistry 28*, 3070−3074.
Rubio, V., Cervera, J., Lusty, C. J., Bendala, E. & Britton, H. G. (1991) *Biochemistry 30*, 1068−1075.
Schott, H. N., & Shetlar, M. D. (1974) *Biochem. Biophys. Res. Commun. 59*, 1112−1116.
Shapiro, A. L., Viñuela, E., & Maizel, J. V. (1967) *Biochem. Biophys. Res. Commun. 28*, 815−820.
Simmer, J. P., Kelly, R. E., Rinker, A. G., Scully, J. L., & Evans, D. R. (1990) *J. Biol. Chem. 265*, 10395−10402.
Stone, K. L., LoPresti, M. B., Crawford, J. M., DeAngelis, R., & Williams, K. R. (1989) in *A Practical Guide to Protein and Peptide Purification for Microsequencing* (Matsudaira, P. T., Ed.) pp 31−47, Academic Press, New York.
Thoden J. B., Raushel, F. M., Mareya, S., Tomchick, D., & Rayment, I. (1995) *Acta Crystallogr. D51*, 827−829.
Yamaguchi, H., Kato, H., Hata, Y., Nishioka, T., Kimura, A., Oda, J., & Katsube, Y. (1993) *J. Mol. Biol. 229*, 1083−1100.

BI952549U