

Amino Acid Sequence of *Escherichia coli* Citrate Synthase[†]

Vipin Bhayana and Harry W. Duckworth*

ABSTRACT: Detailed evidence for the amino acid sequence of allosteric citrate synthase from *Escherichia coli* is presented. The evidence confirms all but 11 of the residues inferred from the sequence of the gene as reported previously [Ner, S. S., Bhayana, V., Bell, A. W., Giles, I. G., Duckworth, H. W., & Bloxham, D. P. (1983) *Biochemistry* 22, 5243]; no information has been obtained about 10 of these (residues 101-108 and 217-218), and we find aspartic acid rather than asparagine at position 10. Substantial regions of sequence homology are noted between the *E. coli* enzyme and citrate synthase from pig heart, especially near residues thought to be involved in

the active site. Deletions or insertions must be assumed in a number of places in order to maximize homology. Either of two lysines, at positions 355 and 356, could be formally homologous to the trimethyllysine of pig heart enzyme, but neither of these is methylated. It appears that *E. coli* and pig heart citrate synthases are formed of basically similar subunits but that considerable differences exist, which must explain why the *E. coli* enzyme is hexameric and allosterically inhibited by NADH, while the pig heart enzyme is dimeric and insensitive to that nucleotide.

Organisms can be divided into two great classes on the basis of the properties of their citrate synthases (Weitzman & Jones, 1968; Weitzman & Danson, 1976). Eukaryotes and Gram-positive bacteria have "small" enzymes, of molecular weight 90 000-100 000, which are insensitive to NADH. In Gram-negative bacteria, with few exceptions, the enzymes are "large", of molecular weights 240 000-280 000, and inhibited specifically by NADH by an allosteric mechanism (Weitzman, 1966; Weitzman & Dunmore, 1969; Tong & Duckworth, 1975; Morse & Duckworth, 1980, and references cited therein). It has been suggested that both kinds of citrate synthase may be based on the same type of subunit, of about 45 000-50 000, and that differences in the subunit interactions account for the presence or absence of allosteric properties (Morse & Duckworth, 1980). Partial proteolysis studies upon citrate synthases from pig heart, a eukaryotic tissue (Bloxham et al., 1980), and from *Escherichia coli*, a Gram-negative bacterium (Bell et al., 1983), suggest that at least some features of the three-dimensional folding of the subunits are the same in both enzymes. The recent publication of the complete amino acid sequences of the pig heart (Bloxham et al., 1981, 1982) and *E. coli* enzymes (Ner et al., 1983) has demonstrated that the subunits indeed show many amino acid homologies.

In the case of the *E. coli* enzyme, the base sequence of the citrate synthase gene was determined, and the statement was made that protein chemistry techniques had confirmed almost all of the amino acid sequence inferred (Ner et al., 1983). In this paper, we present the details of the sequence determination from the protein. We also offer a complete alignment of the sequence with that of pig heart citrate synthase and make some observations on the homologies and differences found.

Materials and Methods

Enzymes and Reagents. Pure *E. coli* citrate synthase was prepared as previously described (Tong & Duckworth, 1975). TPCK-trypsin¹ and soybean trypsin inhibitor were from Sigma. *Lysobacter enzymogenes* endoproteinase was from Boehringer. Iodo[2-¹⁴C]acetic acid was from Amersham. Cyanogen bromide and citraconic anhydride were from Aldrich. Hydroxylamine hydrochloride was from Fisher. *o*-

Iodosobenzoic acid and TFA were from Pierce. HPLC-grade solvents were from Fisher (water, acetonitrile, 2-propanol) or Burdick and Jackson (1-propanol). Chemicals used for automated Edman degradation were from Pierce, except for butyl chloride for which the Beckman product was sometimes used. Pyridine and *N*-ethylmorpholine were refluxed with ninhydrin and distilled before use. Sources of other chemicals have been given previously (Tong & Duckworth, 1975) or were reagent grade.

Methods for Generation of Peptides. Citrate synthase was reduced and carboxymethylated with iodo[2-¹⁴C]acetic acid by the method previously described for S-carboxymethylation of subtilisin-digested enzyme (Bell et al., 1983). For production of arginine peptides, S-carboxymethylated protein was citraconylated by the method of Atassi & Habeeb (1972) as modified by Weng et al. (1978), and the product was dialyzed against 0.05 M *N*-ethylmorpholine and freeze-dried. The dry residue was dissolved in 29 mL of 0.1 M NH₄HCO₃ buffer at a final concentration of 6.0 mg/mL, 0.01 weight of TPCK-trypsin was added, and digestion was allowed to proceed for 3 h at 37 °C. Then a further 0.1 weight of TPCK-trypsin was added, and the digestion continued for another 5 h at 37 °C. The mixture was then treated with a weight of soybean trypsin inhibitor equal to the total TPCK-trypsin used and freeze-dried. For production of methionine peptides, the cyanogen bromide procedure of Steers et al. (1965) was used. Hydroxylamine cleavage of S-carboxymethylated protein was performed with reagent prepared according to Bornstein (1970), except that 6 M Gdn-HCl was also included. Reactions were allowed to proceed 2-4 h at 37 °C and then were stopped by addition of glacial acetic acid, followed by dialysis against water and freeze-drying. In the case of the two attempted hydroxylamine cleavages of peptide CB-1 (see Results), Gdn-HCl was included in one experiment but not the other. In both cases, aliquots of the reaction mixtures were loaded directly onto the HPLC reverse-phase column and eluted with the 1-propanol system (see below). Cleavage at tryptophans with *o*-iodosobenzoic acid was according to Ma-

[†] From the Department of Chemistry, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2. Received October 28, 1983. This work was supported by operating grants from the Natural Sciences and Engineering Research Council of Canada and the University of Manitoba Research Board.

¹ Abbreviations: CNBr, cyanogen bromide; HPLC, high-pressure liquid chromatography; TFA, trifluoroacetic acid; TPCK-trypsin, trypsin pretreated with L-1-(tosylamido)-2-phenylethyl chloromethyl ketone; Gdn-HCl, guanidine hydrochloride; SDS, sodium dodecyl sulfate; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride; EDTA, ethylenediaminetetraacetic acid; PTH, phenylthiohydantoin.

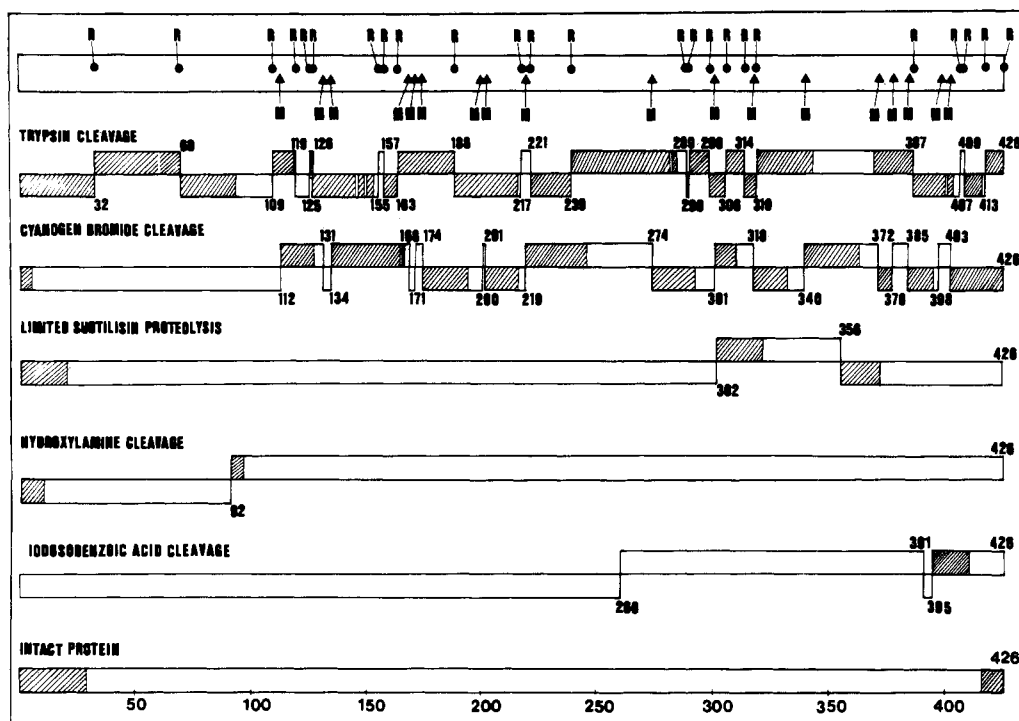


FIGURE 1: Summary of strategy used in sequence determination. The top row locates the arginine (R, ●) and methionine (M, ▲) residues, and the other rows show peptides expected from the indicated cleavage methods; shaded portions were those actually sequenced by automated Edman degradation. In the last row, the shading at the C-terminus shows the information obtained by carboxypeptidase digestion.

honey & Hermodson (1979) as modified by Mahoney et al. (1981) to protect tyrosines. Digestion of peptide CB-1 with *L. enzymogenes* endoproteinase was performed in 0.02 M Tris-HCl buffer, pH 7.8, containing 1 mM EDTA, for 6 h at 37 °C, at a substrate concentration of 1 mg/mL and a proteinase concentration of 0.05 mg/mL.

HPLC Purification of Peptides. All HPLC separations employed the same Perkin-Elmer analytical C8 reverse-phase column, catalog no. 0258-1684. A Perkin-Elmer Series 4 liquid chromatograph was used. Samples were loaded manually, usually as solutions in 6 M Gdn-HCl, and the 0.1% TFA-organic solvent elution systems of Mahoney & Hermodson (1980) were used. Acetonitrile, 2-propanol, and 1-propanol were all used in different experiments; effluents were monitored at 210 nm, and fractions were collected manually as peaks appeared on the recorder.

Other Methods. DEAE-cellulose chromatography was performed in NH_4HCO_3 buffer, and the peptides were eluted by increasing the buffer concentration linearly from 0.03 to 0.5 M. Ion-exchange chromatography of small peptides on Dowex resins used the buffer systems recommended by Schroeder (1972); peptides were located by the ninhydrin method of Moore & Stein (1954). Paper electrophoresis was performed according to the general procedures of Perham (1978), using the volatile pH 6.5 and 2.1 buffers. Procedures for amino acid analysis and automated Edman degradation have already been described (Bell et al., 1983).

Results

Strategy of Sequence Determination. Figure 1 summarizes the sequencing strategy. Most of the information was obtained from two sets of peptides: a set arising from tryptic digestion of citraconylated citrate synthase, named TC-1 through TC-23 in the order in which they appear in the sequence, and a set arising from cleavage of the protein with CNBr, designated CB-1 through CB-19. Additional information was obtained from chemical cleavages at tryptophan with *o*-iodosobenzoic acid and at the unique Asn-Gly bond with hydroxylamine.

The N-terminal sequence of the intact protein and those of two fragments, arising from partial proteolysis of native citrate synthase with subtilisin, have been published previously (Duckworth & Bell, 1982; Bell et al., 1983). Early in the project, the complete sequence of pig heart citrate synthase became available (Bloxham et al., 1981, 1982), and partial homologies between the two enzymes (see Discussion) helped us in establishing a tentative order for many peptides. The recently completed base sequence of the *E. coli* citrate synthase gene (Ner et al., 1983) has allowed detailed checking of all parts of our sequence and gives direct evidence for the order of certain peptides where we did not obtain overlaps. The gene sequence is also the only evidence for one stretch of eight residues (101–108) for which a protein sequence was not obtained in this work.

Arginine Peptides. The mixture obtained from tryptic digestion of 3.1 μmol of S-carboxymethylated, citraconylated citrate synthase was fractionated on Sephadex G-50 in 0.1 M NH_4HCO_3 (Figure 2). Of the six pools obtained, the first proved to contain partially digested material and was not examined further. Fraction II was purified by reverse-phase HPLC using the 1-propanol system (see Materials and Methods) to yield peptide TC-20. Fraction III was separated into four peaks by chromatography on DEAE-cellulose, and each peak was then fractionated by reverse-phase HPLC using the 2-propanol system to yield TC-1, TC-2, TC-3, and TC-14. Fraction IV gave five peaks on DEAE-cellulose, which were fractionated further by reverse-phase HPLC (2-propanol system) to yield TC-7, TC-10, TC-11, and TC-13, and also small amounts of fragments of TC-10 and TC-11, which had arisen by chymotryptic-like cleavages on the C-terminal sides of Tyr-178 and Phe-202, respectively (see below). Fraction V was chromatographed on Dowex 50-X2, and the peaks obtained were purified further by anion exchange on Dowex 1-X2 or paper electrophoresis to give TC-16 and TC-23. Fraction VI was fractionated on Dowex 50-X2, and then combinations of chromatography on Dowex 1-X2, gel filtration in Bio-Gel P2, and paper electrophoresis were used to isolate

Table I: Compositions and Sequencing of Peptides from Tryptic Digestion of Citraconylated Citrate Synthase

peptide: residues covered:	TC-1 ^a (1-32)	TC-2 (33-69)	TC-3 (70-109)	TC-4 (110-119)	TC-5 (120-125)	TC-7 (127-155)	TC-9 (158-163)	TC-10 (164-188)	TC-11 (189-217)
Asp	6.1 (6)	3.4 (3)	4.9 (5)			5.1 (5)		1.0 (0)	4.9 (5)
Thr	4.4 (4)	5.1 (5)	4.5 (5)	2.0 (2)		0.9 (1)		1.5 (1)	1.4 (1)
Ser		3.6 (4)	1.3 (1)			1.6 (2)		1.9 (2)	1.8 (2)
Glu	2.3 (2)	2.4 (2)	6.6 (7)	2.4 (2)		0.4 (0)	1.3 (1)	1.5 (1)	3.2 (3)
Pro		1.3 (1)	1.8 (2)			2.1 (2)		2.8 (3)	2.6 (3)
Gly	3.3 (3)	4.8 (5)	2.6 (2)			1.8 (2)		1.5 (1)	1.7 (1)
Ala	3.3 (3)	1.6 (1)	2.1 (1)	0.7 (0)	1.0 (1)	3.8 (4)	2.0 (2)	2.2 (2)	1.9 (1)
Val	2.7 (3)	1.7 (1)	1.7 (2)			1.8 (2)		1.1 (1)	1.3 (1)
Met		0.4 (0)		0.6 (1)		1.8 (2)		2.2 (3)	1.5 (2)
Ile	1.6 (2)	2.9 (3)	1.8 (2)	1.7 (2)		1.5 (1)	1.1 (1)	1.3 (1)	1.5 (1)
Leu	4.9 (5)	3.0 (3)	4.5 (4)		1.0 (1)	2.4 (2)		2.1 (2)	2.8 (3)
Tyr			2.1 (3)			0.8 (1)		2.5 (3)	1.7 (2)
Phe		3.9 (4)	1.6 (2)		1.9 (2)	0.9 (1)	0.8 (1)	1.2 (1)	2.0 (2)
His	0.2 (0)	0.8 (1)	2.4 (2)	1.6 (2)	1.1 (1)	1.7 (2)			
Lys	3.2 (3)	2.0 (2)	2.3 (2)					1.8 (2)	
Arg	1.0 (1)	0.9 (1)	1.0 (1)	1.0 (1)	1.0 (1)	0.9 (1)	0.8 (1)	1.0 (1)	1.0 (1)
CM-Cys		1.0 (1)	+ ^b (1)			+ (1)		+ (1)	+ (1)
yield ^c (%)	68	85	54	90	36	17 ^c	86	56	60
no. sequenced	32	37	24	9	NS ^e	25	6	25	28
no. unsequenced			94-109	119	120-125	146, 150, 154, 155			217

^a Numbers of residues found by amino acid analysis. In parentheses are numbers expected from the sequence (Figure 4). ^b Present but not quantitated. ^c Calculated from the one-step HPLC purification, except for TC-7, for which the overall yield through the large-scale

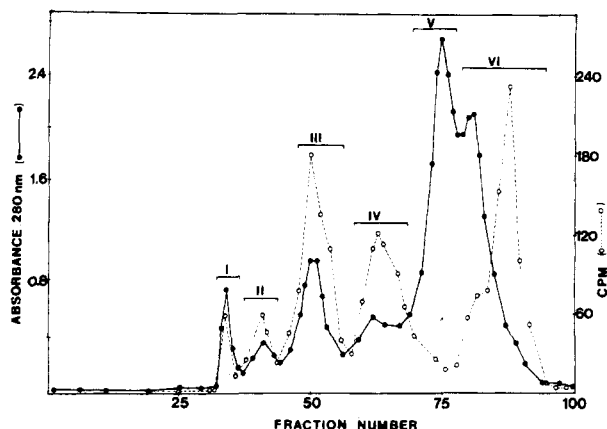


FIGURE 2: Gel filtration of peptides produced by tryptic digestion of S-[¹⁴C]carboxymethylated citraconylated citrate synthase (3.1 μmol). Separation was on a Sephadex G-50 column (2.5 × 160 cm) in 0.1 M NH₄HCO₃. Fractions of 10 mL were collected and monitored for A₂₈₀ (●) and radioactivity in 20-μL aliquots (○). The solid bars indicate the six pools made.

TC-9, TC-18, TC-19, and TC-22. Free arginine was also found in this fraction.

A second batch of arginine peptides was resolved directly by reverse-phase HPLC, by elution with an acetonitrile (Figure 3) or 1-propanol gradient. Recoveries from these columns were used to calculate yields of individual peptides, and in a few cases peptides obtained by this procedure were used for sequencing when material from the first, larger batch had been exhausted. In the case of peptide TC-22, which has N-terminal glutamine, the peptide had become blocked, presumably by formation of pyroglutamic acid, during the lengthy isolation procedure. TC-22 was isolated in unblocked form, however, by the brief, one-step method used for the second batch. Peptides TC-4 and TC-21 were also obtained from the one-step HPLC separation but were not located in the fractions from the first batch.

Details of the compositions, recoveries, and sequencing of the arginine peptides are given in Table I. From the composition there are 24 arginines (Duckworth & Bell, 1982) of which one is C-terminal and another, Arg-407, is in a trypt-

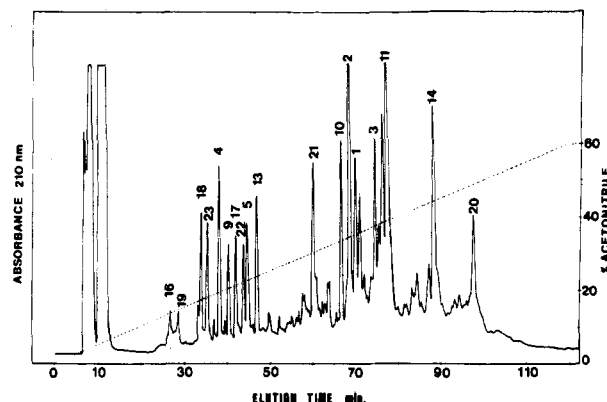


FIGURE 3: Reverse-phase HPLC of tryptic digest of S-[¹⁴C]-carboxymethylated citraconylated citrate synthase. The digest (268 nmol) was decitraconylated and dissolved in 1.0 mL of 6 M Gdn-HCl, and 200 μL was applied to a Perkin-Elmer reverse-phase C8 column (4 × 250 mm) equilibrated with 0.1% TFA. Peptides were eluted with a linear gradient between equilibration solvent and 60% aqueous acetonitrile containing 0.1% TFA over a period of 120 min and flow rate of 1.0 mL/min. The profile shows A₂₁₀ (full scale = 2.56A). The numbers above the peaks refer to the TC peptides subsequently identified in those peaks.

sin-resistant Arg-Pro bond. Twenty-three peptides are therefore expected from the sequence; 19 are listed in Table I. Of the other four, two are free arginine and one is a dipeptide. Our failure to find TC-12 (residues 218-221) means that a crucial overlap of peptides CB-9 and CB-10 is missing from our evidence.

In two cases (TC-1, TC-20), arginine peptides were decitraconylated and digested with trypsin, and the hydrolysis products were separated by reverse-phase HPLC using the 2-propanol and 1-propanol systems, respectively. The peptides from the C-terminal portions of these arginine peptides were completely sequenced; the C-terminal part of TC-20, residues 371-386, includes our only evidence for the sequence of residues 379-385, since the cyanogen bromide peptide expected here was not located (see below). Peptide TC-2 was also decitraconylated and digested with trypsin and the resulting mixture sequenced directly. The short peptide from residues 33-37 washed out within three cycles, leaving two sequences

TC-13 (222-239)	TC-14 (240-289)	TC-16 (291-299)	TC-17 (300-306)	TC-18 (307-314)	TC-19 (315-319)	TC-20 (320-387)	TC-21 (388-409)	TC-22 (410-419)	TC-23 (420-426)
2.0 (2)	2.2 (2)	3.1 (3)		2.1 (2)		7.0 (7)	1.3 (1)		1.8 (2)
2.0 (2)	0.9 (1)				1.3 (1)	3.0 (3)	1.0 (1)	0.7 (1)	
1.9 (2)	4.2 (5)	0.9 (1)				3.4 (3)	2.0 (2)		0.8 (1)
2.0 (2)	4.0 (4)					6.4 (7)	1.3 (1)	1.7 (2)	0.3 (0)
	3.3 (3)			0.9 (1)		3.8 (3)	1.2 (1)		
0.3 (0)	6.0 (6)		2.0 (2)			3.2 (3)	2.0 (2)	1.0 (1)	
2.3 (2)	9.5 (10)	1.4 (1)			1.4 (1)	5.2 (5)	1.9 (2)		
1.1 (1)	2.1 (1)			0.9 (1)	1.1 (1)	3.8 (4)	1.1 (1)		
	1.3 (1)		1.0 (1)		0.9 (1)	4.0 (4)	1.7 (2)		
2.1 (2)	4.2 (4)		0.4 (0)			5.5 (6)	1.9 (2)		1.0 (1)
2.1 (2)	3.3 (3)		1.1 (1)			7.0 (8)		1.1 (1)	
				1.8 (2)		3.1 (3)		1.6 (2)	
	2.3 (3)	0.8 (1)	1.1 (1)			3.0 (4)			1.1 (1)
1.7 (2)	1.8 (2)		1.0 (1)			1.3 (1)	1.6 (2)		
0.3 (0)	1.8 (2)	1.9 (2)		0.9 (1)		4.1 (5)	1.1 (1)	1.1 (1)	1.7 (2)
1.0 (1)	1.0 (1)	1.3 (1)	1.2 (1)	1.0 (1)	1.0 (1)	1.2 (1)	1.5 (1)	1.0 (1)	1.0 (1)
	+					+			
75	63	39	73	90	59	45	86	58	99
18	45	9	7	8	5	42 ^d	14	9	8
	283, 286-289					345-370	400, 403-409		

purification is given (see the text). ^d Residues 371-387 were obtained by sequencing of the C-terminal tryptic peptide derived from TC-20. See the text. ^e NS = not sequenced.

that could be sorted into peptides since one (residues 38-55) was known completely from the sequencing of intact TC-2.

In the case of TC-11, a minor peptide, isolated in low yield from the first, large batch of arginine peptides, was sequenced and proved to have arisen by chymotryptic-like cleavage between Phe-202 and Ser-203. It gave the C-terminal portion of that peptide. A second example of this chymotryptic-like cleavage was also shown by the isolation, in low yield, of a peptide corresponding to the C-terminal part of TC-10 and arising by cleavage between Tyr-178 and Ser-179.

Methionine Peptides. A single batch of the peptides obtained from cyanogen bromide cleavage of 0.44 μ mol of S-carboxymethylated citrate synthase was fractionated directly by reverse-phase HPLC using the 2-propanol system. Most peaks contained single peptides, sometimes with 10-20% contamination by their neighbors as recognized by sequencing. CB-9 and CB-13, which were poorly separated on this column, were resolved by rechromatography on the same column using the 1-propanol system. Peptides CB-11 and CB-17, which were found in a single peak, could not be resolved by rechromatography and were sequenced together; rapid washout of CB-17, after the removal of arginine at cycle 2, led to a de facto pure sequence for CB-11 after a few cycles.

Of the 19 peptides expected from the sequence, 13 were isolated. Of the other six, one is free homoserine and three are tripeptides, which were not found. The remaining two missing peptides were of five (residues 399-403) and seven (residues 379-385) amino acids; they could not be located in the fraction from the HPLC run, but a six-residue peptide, CB-15, was isolated in good yield although it is not clearly more hydrophobic than the two missing peptides. Other methods evidently would have been needed to ensure dependable isolation of all short peptides.

Because of its size, 112 residues, and because its N-terminal sequence was already available from sequenator runs on peptide TC-1 and the intact protein, CB-1 was sequenced for only six cycles to confirm its identity. To obtain sequence information from the C-terminal portion of this peptide, we attempted to cleave CB-1 at its Asn-Gly sequence, residues 91 and 92, with hydroxylamine. Although this reaction was successful when it was used on the intact protein (see below),

no cleavage of CB-1 was detected, by reverse-phase HPLC of the reaction mixture, in two attempts. Perhaps the acidic conditions used in the original cyanogen bromide cleavage, or some aspect of the subsequent peptide purification, led to complete deamidation of Asn-91. We also digested CN-1 with *L. enzymogenes* endoprotease "Lys-C" and separated the hydrolysis products by reverse-phase HPLC. Peptides having the compositions for the larger expected products, including residues 7-21, 22-37, 38-55, and 56-104, were readily identified, but the smaller peptides, including that covering the unsequenced region 105-112, were not found and perhaps did not bind to the column under our conditions. Since residues 56-104 were found in a single peptide, the enzyme evidently did not cleave the Lys-Pro bond at residues 94 and 95.

Hydroxylamine Cleavage. The total products from hydroxylamine cleavage of citrate synthase were run in the sequenator, and two sequences only were found—the N-terminal sequence of the protein and one commencing with glycine and corresponding to the sequence from Gly-92 on. The yields of alanine and glycine in cycle 1, starting with 78 nmol of protein, were 83 and 26 nmol, indicating a cleavage yield of about 33%. The cleavage products were fractionated by gel filtration on Sephadex G-75, in the presence of 6 M Gdn-HCl, and two peaks were obtained. The smaller one could be identified from composition and sequence as residues 1-91, but the larger peak was a mixture, partially resolved, of the other cleavage product (residues 92-426) with uncleaved material. Several fractions from different parts of this large peak were analyzed by SDS-polyacrylamide gel electrophoresis, and three, which seemed to have the highest proportion of cleavage product, were sequenced. All still showed the N-terminal sequence of intact protein, but the second sequence could be assigned unambiguously up to cycle 9 (Tyr-100). It should be noted that no evidence was found for cleavage between residues 10 and 11 of the protein, which are Asp-Gly according to our data but were predicted to be Asn-Gly by the gene sequence of Ner et al. (1983).

Oxidative Cleavage at Tryptophan. The three tryptophans of *E. coli* citrate synthase are located at residues 260, 391, and 395, and the only product likely to yield new sequence information, after compilation of evidence from the data al-

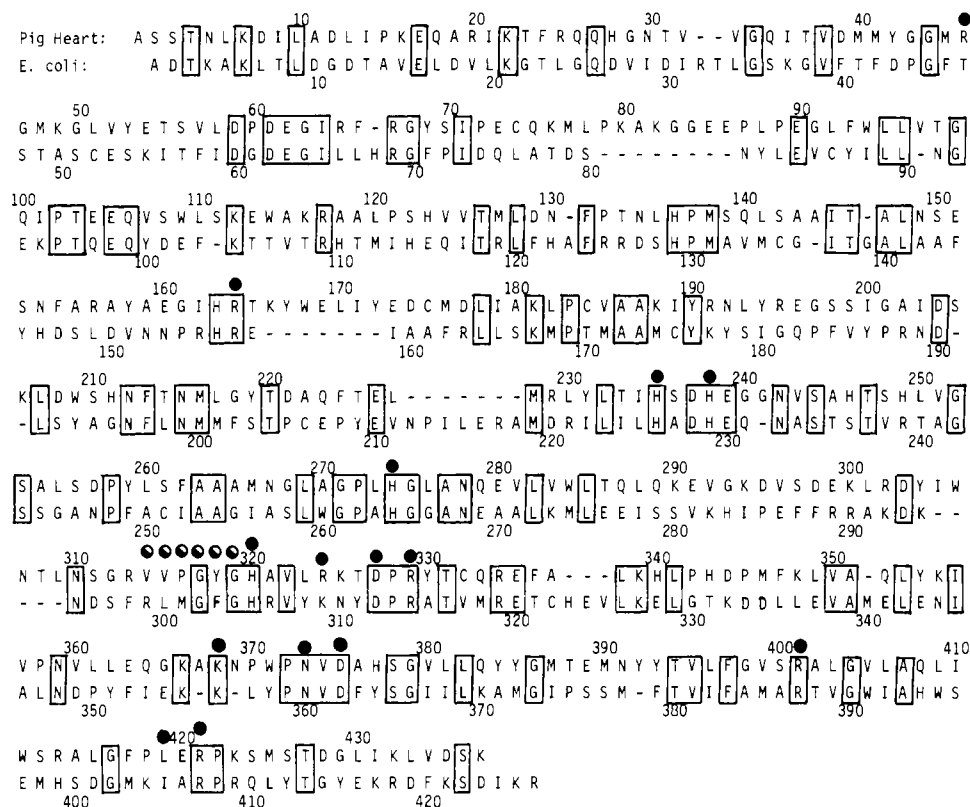


FIGURE 4: Alignment of amino acid sequences of pig heart (top) and *E. coli* (bottom) citrate synthases. Gaps have been introduced in both sequences to maximize homologies and are shown by dashes. Boxes are drawn around identities, and black dots (●) have been placed above residues apparently involved in catalysis or substrate binding, according to the X-ray structure of pig heart enzyme published by Remington et al. (1982). Residues marked (●) in the pig heart sequence, according to the same authors, provide peptide backbone atoms but not side chains to the CoA binding site.

ready presented, was that arising from cleavage at Trp-395. Accordingly, the products of cleavage of citrate synthase with *o*-iodosobenzoic acid were fractionated on Sephadex G-50 in 9% formic acid, and the last to elute of the three A_{280} peaks was chromatographed on the reverse-phase HPLC column using the acetonitrile system. One of two major peaks from this fractionation was sequenced and proved to be the peptide required commencing with Ser-396. Sixteen cycles of automatic Edman degradation were run on this peptide, giving the expected amino acid in each case; an unusual PTH-amino acid, found in cycles 3 and 8 of this degradation, had the elution time of PTH-methionine sulfone, as expected from the oxidation conditions used to perform the cleavage.

Carboxypeptidase Digestion of Intact Citrate Synthase. A sample of S-carboxymethylated citrate synthase, 106 nmol, was digested with 0.01 weight of both carboxypeptidases A and B. Amino acids found after different digestion times are listed in Table II. The amounts, and order of appearance, are consistent with the belief that TC-23 and CB-19 are from the C-terminus of the protein.

Discussion

Assembly of the Sequence. The evidence in this paper, taken by itself, is not quite sufficient to prove the full sequence (shown in Figure 4), but it confirms all but 11 of the residues inferred from the sequence of the gene as reported by Ner et al. (1983). There is no reason to doubt the few residues not confirmed here. It is convenient to discuss the proof of sequence in four sections.

Residues 1-109. The N-terminal sequence of the intact protein (Duckworth & Bell, 1982) allowed the initiation codon in the gene to be recognized, and two laboratories have published DNA sequences that agree up to the codon for Leu-90

Table II: Amino Acids Released by Carboxypeptidases A and B from 106 nmol of S-Carboxymethylated Citrate Synthase

amino acid	no. of nmol after digestion times			
	10 min	30 min	60 min	240 min
arginine	117	122	164	240
lysine	70	130	193	253
isoleucine	17	57	70	81
aspartic acid		50	95	144
serine		28	95	131
phenylalanine		20	58	75

sequence deduced:
 -(Lys,Arg,Asp,Phe,Lys)Ser-Asp-Ile-Lys-Arg-CO₂H

(Ner et al., 1983; Hull et al., 1983). We have placed peptides TC-1, -2, -3 in that order by reference to these DNA sequences and have not sought further evidence to prove it. It may be noted in passing, however, that the compositions of the peptides derived from lysine-specific digestion of CB-1 were consistent with the resulting sequence. The sequence we obtained for TC-3 extends to Lys-94 and overlaps the hydroxylamine cleavage product, whose sequence extends to Tyr-100 as already noted. We failed to obtain any sequence information for residues 101-108, and the DNA sequence is thus the only evidence for this stretch. For what it is worth, the composition of TC-3 (see Table I), which covers this section, is consistent with the proposed sequence, which is rich in threonine.

Residues 110-219. We have continuous sequence information all the way through this region as far as Glu-216. The composition of CB-9 is consistent with the sequence inferred from the DNA. Since we did not find the short peptide TC-12, which would comprise residues 218-221 in the full sequence, we have no evidence from protein chemistry that peptide CB-10 follows CB-9 directly.

Residues 220–299. The sequence of these residues may be assembled from the relevant peptides with no ambiguities. Since we were unable to sequence CB-11 beyond cycle 20, there is no proof from our work that TC-17 follows directly upon TC-16.

Residues 300–426. This stretch is mostly proved by relevant arginine and methionine peptides. To prove the stretch between 364 and 387, it is necessary also to use the sequence of the C-terminal tryptic peptide from TC-20 and the N-terminal sequence of the subtilisin partial proteolysis fragment, commencing at Lys-356, which was published previously (Bell et al., 1983). Proof of the sequence of residues 399–404 is provided by the peptide isolated after oxidative cleavage at tryptophan.

Comparison with the Sequence of Pig Heart Citrate Synthase. The sequence of one other citrate synthase—that from pig heart—is available in the literature (Bloxham et al., 1981, 1982). The three-dimensional structure of this molecule is also known (Remington et al., 1982), and certain amino acids have been identified as active site residues (Remington et al., 1982). We have previously noted that most of these residues have counterparts in the *E. coli* sequence (Bell et al., 1983; Ner et al., 1983). A detailed alignment of the two sequences is shown in Figure 4. The total number of identities shown is 117, but several gaps or insertions have been assumed, in order to make the homology better. It can be seen that regions of apparently significant homology are scattered throughout the sequence and that they do not always correspond to areas believed to be involved in the active site. Although two lysines are potentially homologous to the trimethyllysine of pig heart enzyme, we found no evidence that either of them is methylated.

There are also considerable stretches that show no homology. There appears, in particular, to be little or no conservation of residues identified by Remington et al. (1982) as involved in the intersubunit contact in dimeric pig heart citrate synthase (details of comparison not shown). In the case of the *E. coli* enzyme, a larger, hexameric structure is found (Tong & Duckworth, 1975), and there is an efficient and highly specific site for the allosteric inhibitor NADH, which distinguishes clearly the reduced from the oxidized form of that nucleotide (Duckworth & Tong, 1976). In view of these substantial differences in quaternary structure and functional properties, it is not surprising that these two citrate synthases should show great divergence in sequence. The homologies, on the other hand, are in keeping with the expectation that citrate synthase is an ancient enzyme whose basic catalytic structure should be conserved as evolution proceeds.

Acknowledgments

We thank Isamu Suzuki for unrestricted access to the Sharples centrifuge and French press and Alexander Bell for discussions.

Registry No. Citrate synthase, 9027-96-7.

References

- Atassi, M. Z., & Habeeb, A. F. S. A. (1972) *Methods Enzymol.* 25, 546–553.
- Bell, A. W., Bhayana, V., & Duckworth, H. W. (1983) *Biochemistry* 22, 3400–3405.
- Bloxham, D. P., Ericsson, L. H., Titani, K., Walsh, K. A., & Neurath, H. (1980) *Biochemistry* 19, 3979–3985.
- Bloxham, D. P., Parmelee, D. C., Kumar, S., Wade, R. D., Ericsson, L. H., Neurath, H., Walsh, K. A., & Titani, K. (1981) *Proc. Natl. Acad. Sci. U.S.A.* 78, 5381–5385.
- Bloxham, D. P., Parmelee, D. C., Kumar, S., Walsh, K. A., & Titani, K. (1982) *Biochemistry* 21, 2028–2036.
- Bornstein, P. (1970) *Biochemistry* 9, 2408–2421.
- Duckworth, H. W., & Tong, E. K. (1976) *Biochemistry* 15, 108–114.
- Duckworth, H. W., & Bell, A. W. (1982) *Can. J. Biochem.* 60, 1143–1147.
- Hull, E. P., Spencer, M. E., Wood, D., & Guest, J. R. (1983) *FEBS Lett.* 156, 366–369.
- Mahoney, W. C., & Hermodson, M. A. (1979) *Biochemistry* 18, 3810–3814.
- Mahoney, W. C., & Hermodson, M. A. (1980) *J. Biol. Chem.* 255, 11199–11203.
- Mahoney, W. C., Smith, P. K., & Hermodson, M. A. (1981) *Biochemistry* 20, 443–448.
- Moore, S., & Stein, W. H. (1954) *J. Biol. Chem.* 211, 907–913.
- Morse, D., & Duckworth, H. W. (1980) *Can. J. Biochem.* 58, 696–706.
- Ner, S. S., Bhayana, V., Bell, A. W., Giles, I. G., Duckworth, H. W., & Bloxham, D. P. (1983) *Biochemistry* 22, 5243–5249.
- Perham, R. N. (1978) in *Techniques in Protein and Enzyme Biochemistry*, Part 1, pp 1–39, Elsevier, Amsterdam.
- Remington, S., Wiegand, G., & Huber, R. (1982) *J. Mol. Biol.* 158, 111–152.
- Schroeder, W. A. (1972) *Methods Enzymol.* 25, 203–221.
- Steers, E., Jr., Craven, G. R., Anfinsen, C. B., & Bethune, J. B. (1965) *J. Biol. Chem.* 240, 2478–2484.
- Tong, E. K., & Duckworth, H. W. (1975) *Biochemistry* 14, 235–241.
- Weitzman, P. D. J. (1966) *Biochim. Biophys. Acta* 128, 213.
- Weitzman, P. D. J., & Jones, D. (1968) *Nature (London)* 219, 270–272.
- Weitzman, P. D. J., & Dunmore, P. (1969) *Biochim. Biophys. Acta* 171, 198–200.
- Weitzman, P. D. J., & Danson, M. J. (1976) *Curr. Top. Cell. Regul.* 10, 161–204.
- Weng, L., Russell, J., & Heinrikson, R. L. (1978) *J. Biol. Chem.* 253, 8093–8101.