(Ho, C., Ed.) Elsevier/North-Holland, New York (in press).
Lyons, K. B., Friedman, J. M., & Fleury, P. A. (1976) *Nature (London) 275*, 565-566.
Makinen, M. W. (1975) in *Techniques and Topics in Bio-Inorganic Chemistry* (McAuliffe, C. A., Ed.) pp 39-43, Macmillan, London.
Makinen, M. W., & Eaton, W. A. (1974) *Nature (London) 247*, 62-64.
Moffat, K., Deatherage, J. F., & Seybert, D. W. (1979) *Science (Washington, D.C.) 206*, 1035-1042.
Nagai, K., Welborn, C., Dolphin, D., & Kitagawa, T. (1980) *Biochemistry 19*, 4755-4761.
Noe, L. J., Eisert, W. G., & Rentzepis, P. M. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 573-577.
Ondrias, M. R., Rousseau, D. L., & Simon, S. R. (1981) *Science (Washington, D.C.) 213*, 657-659.
Ondrias, M. R., Rousseau, D. L., & Simon, S. R. (1982) *Biochemistry* (in press).
Perutz, M. F. (1970) *Nature (London) 228*, 726-734.
Perutz, M. F., Kilmartin, J. V., Nagai, K., Szabo, A., & Simon, S. R. (1976) *Biochemistry 15*, 378-387.
Pulsinelli, P. D. (1973) *J. Mol. Biol. 74*, 57-66.
Reynolds, A. H., Rand, S. D., & Rentzepis, P. M. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2292-2296.
Rousseau, D. L. (1981) *J. Raman Spectrosc. 10*, 94-99.
Rousseau, D. L., Shelnutt, J. A., Ondrias, M. R., Friedman, J. M., Henry, E. R., & Simon, S. R. (1981) in *Interactions Between Iron and Proteins in Oxygen and Electron Transport* (Ho, C., Ed.) Elsevier/North-Holland, New York (in press).
Saffron, W. A., & Gibson, Q. H. (1977) *J. Biol. Chem. 252*, 7955-7958.
Sawicki, C. A., & Gibson, Q. H. (1974) *J. Biol. Chem. 251*, 558-562.
Sawicki, C. A., & Gibson, Q. H. J. (1979) *J. Biol. Chem. 254*, 4058-4062.
Scheidt, R. W. (1977) *Acc. Chem. Res. 10*, 339-343.
Shank, C. V., Ippen, E. P., & Bersohn, R. (1976) *Science (Washington, D.C.) 193*, 50-51.
Shelnutt, J. A., Rousseau, D. L., Dethmers, J. K., & Margoliash, E. (1979a) *Proc. Natl. Acad. Sci. U.S.A. 76*, 3865-3869.
Shelnutt, J. A., Rousseau, D. L., Friedman, J. M., & Simon, S. R. (1979b) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4409-4413.
Spaulding, L. D., Chang, C. C., Yu, N.-T., & Felton, R. H. (1975) *J. Am. Chem. Soc. 97*, 2517-2522.
Spiro, T. G. (1981) *Isr. J. Chem. 21*, 81.
Spiro, T. G., & Burke, J. M. (1976) *J. Am. Chem. Soc. 98*, 5482-5489.
Takano, T. (1977) *J. Mol. Biol. 110*, 569-584.
Terner, J., Spiro, T. G., Nagumo, M., Nicol, M. F., & El-Sayed, A. (1980) *J. Am. Chem. Soc. 102*, 3238-3239.
Terner, J., Spiro, T. G., Nagumo, M., Nicol, M. F., & El-Sayed, A. J. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 1313-1317.
Tsubaki, M., Nagai, K., & Kitagawa, T. (1980) *Biochemistry 19*, 379-385.
Warshel, A. (1977) *Annu. Rev. Biophys. Bioeng. 6*, 273-300.
Warshel, A., & Weiss, R. M. (1981) *J. Am. Chem. Soc. 103*, 446-451.

Complete Amino Acid Sequence of Porcine Heart Citrate Synthase†

David P. Bloxham,‡ David C. Parmelee,§ Santosh Kumar, Kenneth A. Walsh, and Koiti Titani*

ABSTRACT: The detailed proof of the 437-residue amino acid sequence (M_r 48 969) of porcine heart citrate synthase (EC 4.1.3.7) is described. The S-carboxymethylated protein has been cleaved at methionine (cyanogen bromide) and arginine (trypsin digest of citraconylated enzyme) residues to yield 14 and 17 major peptides, respectively. Peptides were initially fractionated by gel filtration, and those useful for sequence analysis were purified by high-performance liquid chromatography. Sequence analyses were performed on these primary peptides and on subpeptides generated by cleavage with the bromine adduct of 2-[(2-nitrophenyl)sulfenyl]-3-methylindole, *Staphylococcus aureus* V8 protease, trypsin, chymotrypsin, or acid. The overall sequence was confirmed by analyzing products of cleavage by hydroxylamine, acid, and subtilisin. A novel feature of the sequence is the identification of trimethyllysine at residue 368.

Citrate synthase [citrate oxaloacetate-lyase [coenzyme A (CoA)[1] acetylating]; EC 4.1.3.7] catalyzes the aldol-type condensation of oxaloacetate and acetyl-CoA. The analysis of this enzyme is closely linked to landmarks in the history of biochemical studies such as the discovery of the cyclic nature of the citric acid cycle [reviewed by Ochoa (1954)] and the identification of "active acetate" as acetyl-CoA (Lynen & Reichert, 1951; Lynen et al., 1951; Stern et al., 1951). The enzyme was first isolated in crystalline form from pig heart by Ochoa and his collaborators (Stern et al., 1950; Ochoa et al., 1951) and has subsequently been purified from a wide variety of sources (Weitzman & Danson, 1976). The most common form of the enzyme has a molecular mass of 90 000-100 000 and is made up of two identical subunits (Singh et al., 1970; Srere, 1975). The stereochemical course of the reaction mechanism has been elucidated in a number of elegant studies which have been comprehensively reviewed (Spector, 1972; Srere, 1975). In addition, enzyme activity is

† From the Howard Hughes Medical Institute and the Department of Biochemistry, University of Washington, Seattle, Washington 98195. *Received October 26, 1981.* This work was supported by grants from the National Institutes of Health (GM 15731) and the Wellcome Trust.
‡ Permanent address: Department of Biochemistry, University of Southampton, Southampton, S09 3TU, United Kingdom.
§ Present address: Deutsches Krebsforschungszentrum, D-6900 Heidelberg, West Germany.

[1] Abbreviations: BNPS-skatole, bromine adduct of 2-[(2-nitrophenyl)sulfenyl]-3-methylindole; CM, *S*-carboxymethyl; HPLC, high-performance liquid chromatography; $NaDodSO_4$, sodium dodecyl sulfate; Tos-LysCH_2Cl, L-1-chloro-3-(tosylamido)-7-amino-2-heptanone; Pth, phenylthiohydantoin; CoA, coenzyme A.

0006-2960/82/0421-2028$01.25/0

allosterically regulated by a number of ligands which are important in controlling the citric acid cycle (Srere, 1975; Weitzman & Danson, 1976).

Until recently, little was known abut the structures of citrate synthase or other enzymes which use acetyl-CoA as a substrate. A 3.5-nm resolution X-ray crystal structure was reported (Wiegand et al., 1979) which defined the general shape of the molecule, the high proportion of α-helical content, and the number and location of several binding sites. This structure could not identify any of the individual side chains. As an initial step in the process of sequencing citrate synthase, a procedure for limited proteolysis by subtilisin was developed (Bloxham et al., 1980) which in the presence of palmitoyl-CoA specifically cleaved the protein into an amino-terminal portion (CS subt 37) and a carboxyl-terminal segment (CS subt 10). This analysis is now extended to the isolation of appropriate cyanogen bromide, tryptic, and other fragments which have enabled a complete sequence to be deduced. Preliminary reports of this sequence have been presented (Bloxham et al., 1981a,b).

Materials and Methods

Pig heart citrate synthase was either purified by the method of Srere (1969) or purchased from the Sigma Chemical Co.; Tos-LysCH_2Cl-trypsin and chymotrypsin were obtained from Worthington Biochemical Corp. *Staphylococcus aureus* V8 protease was purchased from Miles Laboratories. Sequenal grade chemicals for sequencing, BNPS-skatole, fluorescamine, and trifluoroacetic acid were products of Pierce Chemical Co. The trifluoroacetic acid was refluxed twice over solid CrO_3, dried over anhydrous $CaSO_4$, and fractionally distilled before use. Organic solvents were obtained from Burdick and Jackson. Water and ethanol for HPLC were glass distilled prior to use. Cyanogen bromide was obtained from Eastman. N^{ϵ}-Trimethyl-L-lysine oxalate was a product of Bachem Inc. Fine Chemicals. Guanidine hydrochloride (ultrapure) was purchased from Heico. Sephadex, Sephacryl, and ion-exchange Sephadex of various grades were obtained from Pharmacia Fine Chemicals.

Automated sequence analyses were performed with a Beckman sequencer (model 890) according to Edman & Begg (1967) as modified by either Hermodson et al. (1972) or Brauer et al. (1975) which gave consistently yields of 91–96%. Small peptides were sequenced in the presence of Polybrene (Tarr et al., 1978). Phenylthiohydantoin (Pth) derivatives of the amino acids were identified by two complementary systems of reversed-phase HPLC (Ericsson et al., 1977; Hermann et al., 1978). Amino acid analyses were performed with a Dionex amino acid analyzer (Model D-500). Analyses involving less than 1.5 nmol of peptide tended to be poor, particularly for serine, glycine, and glutamic acid, even with peptides purified by HPLC. Trimethyllysine was determined in acid hydrolysates by the method of Van Eldik et al. (1980).

The conditions for protein carboxymethylation with iodo-[1-^{14}C]acetic acid, cyanogen bromide cleavage, succinylation, trypsin digestion, BNPS-skatole cleavage at tryptophanyl bonds, chymotrypsin digestion, *S. aureus* V8 protease digestion, gel filtration on Sephadex, chromatography on SP Sephadex C-25, and paper electrophoresis were described by Koide et al. (1978) and were used with the following minor modifications. Trypsin digestions generally were applied to the citraconylated (Atassi & Habeeb, 1972) CM protein and were terminated by the addition of lima bean trypsin inhibitor. Citraconyl groups were removed in 9% formic acid (3 h at 37 °C).

Generation of Large Fragments of Citrate Synthase. The conditions for limited proteolysis by subtilisin BPN′ (Teikoku Chemical Corp.) and isolation of fragments CS subt 37 (residues 1–321) and CS subt 10 (residues 322–437) were described previously (Bloxham et al., 1980). Cleavage at aspartyl–proline bonds (Piszkiewicz et al., 1970) was achieved by reacting the protein (2 mg/mL) in 70% formic acid containing 0.1% β-mercaptoethanol for 36 h at 38 °C. The sample was then diluted with 10 volumes of water and freeze-dried. For hydroxylamine cleavage of asparaginyl–glycine bonds (Bornstein, 1970), [^{14}C]CM-citrate synthase (60 mg) was reacted for 7 h at 45 °C in 30 mL of 2 M NH_2OH–6 M guanidine hydrochloride, pH 9. The reaction was stopped by adding 3 mL of 80% formic acid, and the mixture was dialyzed exhaustively vs. water. After being freeze-dried, the protein was dissolved in 5 mL of 6 M guanidine hydrochloride–0.1 M NH_4HCO_3. Fragments HAα (residues 1–267) and HAβ (residues 268–437) were separated by chromatography on a column (2.5 × 100 cm) of Sephacryl S-200 equilibrated in the same buffer at room temperature by using a flow rate of 24 mL/h. Fragments HAα and HAβ corresponded to the first two major protein fractions. Fragment HAα-AC1 (residues 60–257) was isolated by chromatography under identical conditions. Molecular weights of large fragments were determined by NaDodSO_4–polyacrylamide gel electrophoresis (Laemmli, 1970).

Reversed-Phase HPLC. Peptide mixtures were dissolved in small volumes of mobile phase solvent or 6 M guanidine hydrochloride and then chromatographed on one of the following four systems: system 1, μBondapak C_{18} (Waters Instruments), mobile phase 0.1% trifluoroacetic acid, mobile phase modifier ethanol containing 0.07% trifluoroacetic acid; system 2, μBondapak C_{18}, mobile phase 0.1% trifluoroacetic acid, mobile phase modifier acetonitrile containing 0.07% trifluoroacetic acid; system 3, μBondapak CN (Waters Instruments), mobile phase 50 mM phosphoric acid titrated to pH 3 with 8 M NaOH, mobile phase modifier acetonitrile; system 4, μBondapak C_{18}, mobile phase 0.01 M HCl, mobile phase modifier acetonitrile. All chromatography was performed at room temperature, and the flow rate was varied as appropriate. Peptide samples containing organic solvents were first exposed to a vacuum to remove most of the solvent. They were then diluted with water and freeze-dried. When phosphate buffer was used as the mobile phase, peptide fractions were desalted on a Sephadex G-25 column in 0.1% trifluoroacetic acid and then freeze-dried.

Peptide Nomenclature. This follows the convention (cleavage 1) (order in sequence)-(cleavage 2) (order in sequence)- etc., where cleavage reactions are designated as follows: AC, acid; CB, cyanogen bromide; HA, hydroxylamine; T, trypsin; Tc, trypsin at arginine; W, BNPS-skatole.

Results

Strategy. To solve the structure of citrate synthase, we isolated a complete set of cyanogen bromide (CB) cleavage peptides from CM-citrate synthase and sequenced these as completely as possible. The CB peptides were then ordered by isolation and analysis of tryptic cleavage peptides from CM-citraconylated citrate sythase. Certain features of the structure were also confirmed by isolation of large fragments derived by limited proteolysis (Bloxham et al., 1980), hydroxylamine cleavage at asparaginyl–glycine bonds (Bornstein, 1970), and acid cleavage at aspartyl–proline bonds (Piszkiewicz et al., 1970). Although our approach to sequence analysis is conventional, a key feature of the current work is the extensive application of HPLC for the rapid purification of selected peptides.

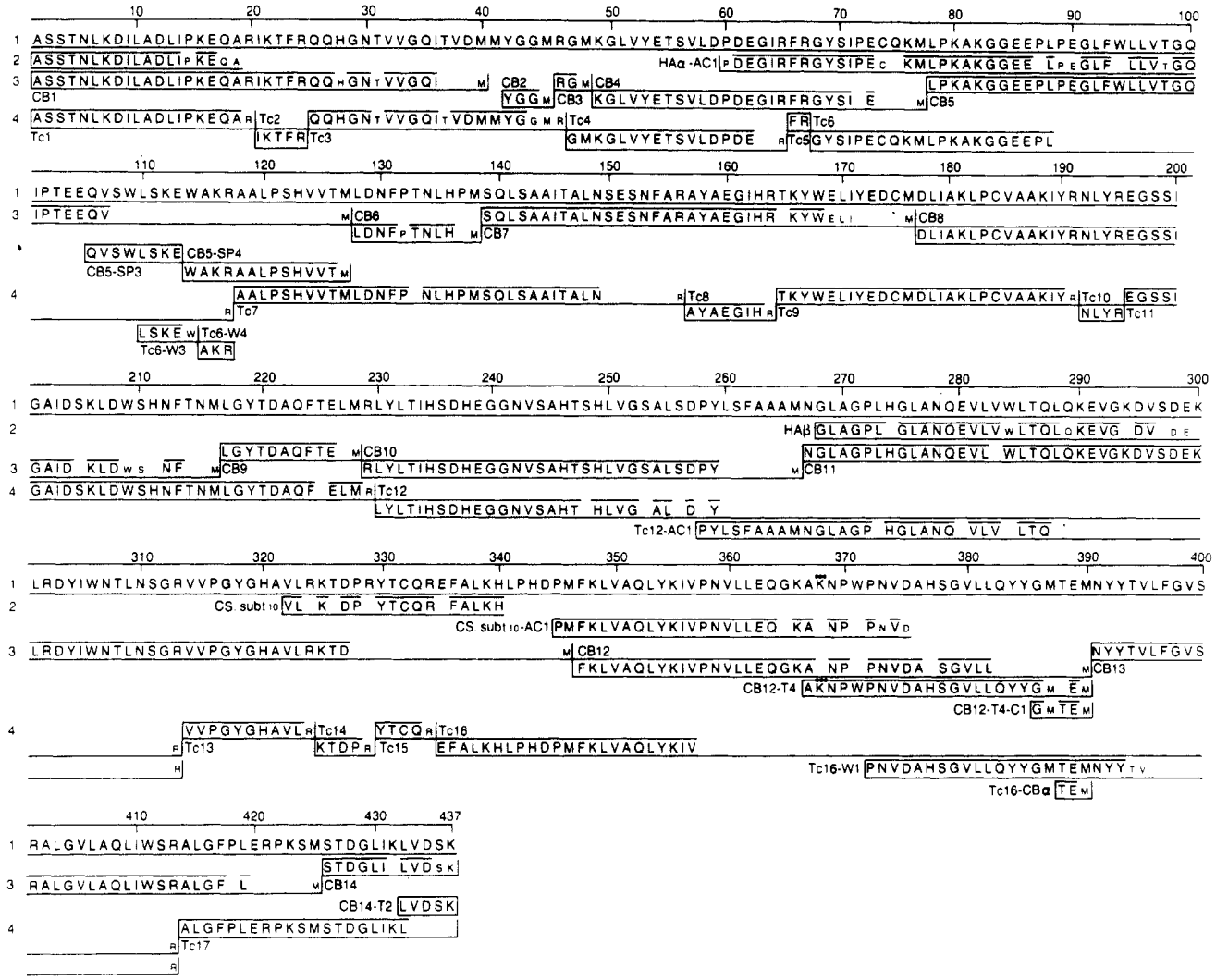

FIGURE 1: Summary proof of the structure of citrate synthase. The one-letter code within the bars designates amino acid residues in that peptide. Large capital letters indicate residues positively identified by Edman degradation. Small capital letters indicate residues which were placed by cleavage specificity or which gave weak identifications in the Edman degradation. The length of each bar indicates the length of the peptide except for the intact protein (1–437), HAα-AC1 (60–267), HAβ (268–437), CS subt 10 (322–437), and CS subt 10-AC1 (345–437); enclosure of the top of the bar indicates that that portion of the sequence is proven; gaps in the upper enclosure indicate portions of the sequence not identified. The numbered lines (left) indicate the following: (1) final sequence; (2) sequence of large protein fragments; (3) sequence of CB peptides; (4) sequence of Tc peptides. One-letter amino acid abbreviations are the following: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; K with three dots, trimethyllysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

In the presentation of the results, we have shown only the primary data involved in isolating and sequencing the CB peptides. All additional information is presented as a supplement as indicated at the end of this paper (see paragraph at end of paper regarding supplementary material).

Sequence of Intact Citrate Synthase. A complete summary proof of the sequence of the 437 residues of citrate synthase is shown in Figure 1. The amino terminus of the enzyme was identified by sequencing 19 residues of the intact CM protein. Part of this sequence was reported previously (Bloxham et al., 1980). Both Pth-Ala and Pth-Ser (ratio 3.6:1) were identified in the first turn of the Edman degradation, indicating that the amino terminus was "ragged". The summed yield of both amino acids was consistent with the absence of any blocking groups on the amino terminus.

Purification of CB Peptides. [^{14}C]CM-citrate synthase was cleaved with cyanogen bromide, and the mixture was resolved into 10 fractions by gel filtration (Figure 2). Most of the radioactivity was found in fractions III, V, and VI. Fractions I and II contained the products of incomplete cleavage reactions and were not analyzed further. The remaining fractions were analyzed by NaDodSO$_4$–15% polyacrylamide gel electrophoresis. Since peptides in fractions VII–X were too small to be retained on the gel, they must contain less than 25 residues. The remainder of the fractions showed a narrow molecular mass distribution varying between 8000 and 10000 (fraction III) and 4000 and 6000 (fractions IV–VI). The similarity of molecular size suggested that further gel filtration would have only limited success. Furthermore, purification of peptides in the intermediate M_r range (4000–6000) by ion-exchange chromatography is frequently rather difficult, although fraction III was chromatographed on SP Sephadex C-25 in the presence of 7 M urea (Figure 3) to yield two radioactive fragments. The amino terminus of the first fraction was Lys-Gly-Leu-Val-, which is identical with that of a smaller peptide, CB4. Since this was an overlap peptide (residues 49–127) formed by incomplete cleavage between CB4 and CB5, it was not analyzed further. CB11 contained a single polypeptide and was suitable for automated sequencing.

In view of the recent dramatic successes in the separation of peptides by HPLC, especially as applied to bacteriorhodopsin (Gerber et al., 1979), we decided to adapt these

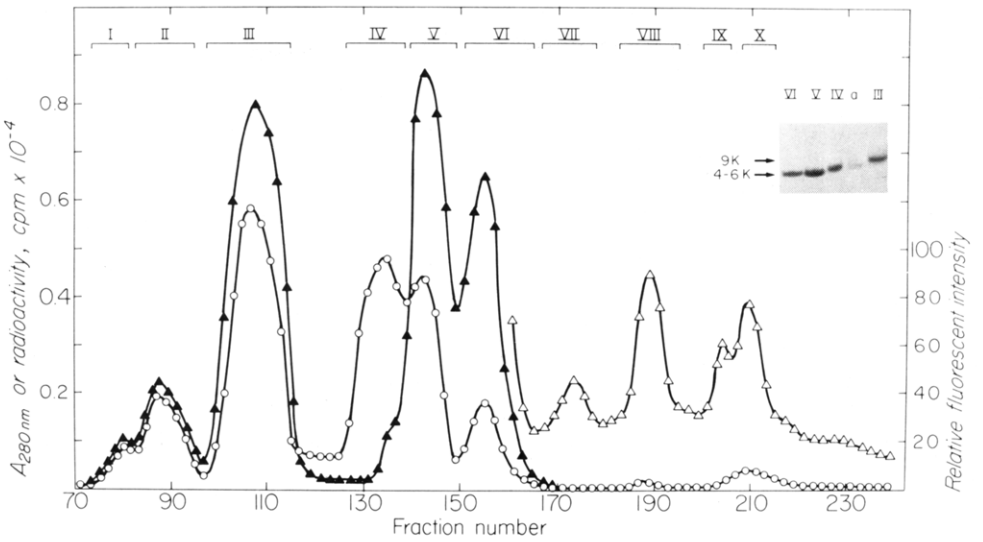

FIGURE 2: Fractionation of cyanogen bromide fragments of [^{14}C]CM-citrate synthase (65 mg) at room temperature at a flow rate of 20 mL/h on a column (2.5 × 200 cm) of Sephadex G-50 superfine in 9% formic acid. The column was monitored for absorbance at 280 nm (○), fluorescence after reaction with fluorescamine (△), and radioactivity per 25-μL sample (▲). Fractions of 5 mL were collected and pooled as indicated. The insert shows a $NaDodSO_4$–15% polyacrylamide gel of the pooled fractions. Track a represents the pooled fractions between III and IV.

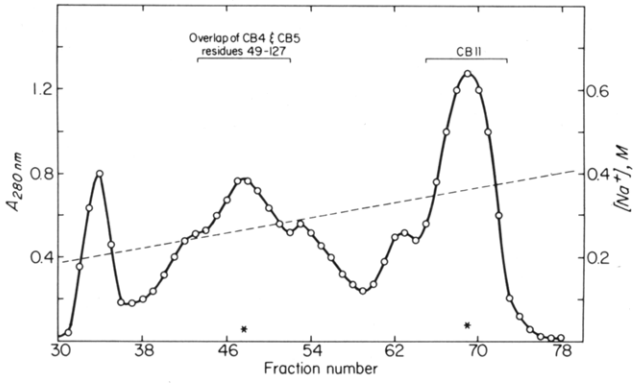

FIGURE 3: Fractionation of 11 mg of cyanogen bromide fraction III on a column (1 × 50 cm) of SP Sephadex C-25. The peptides were eluted at room temperature at a flow rate of 12 mL/h with a 300-mL linear gradient of 0.1 M sodium formate–7 M urea, pH 2.9, to 0.75 M sodium formate–7 M urea, pH 4. Fractions of 2 mL were collected and pooled as indicated. The asterisk indicates the presence of radioactivity.

techniques to the purification of the other peptides from citrate synthase. The resolution of cyanogen bromide fractions IV–VIII is shown in Figure 4. Based on experience gained in the course of this work, we conclude that systems 1 and 2 described under Materials and Methods are optimal. Four pure peptides (CB1, CB5, CB12, and CB13) were obtained from cyanogen bromide fraction IV. CB12 and CB13 were also present in fraction V but in somewhat lower yield. Cyanogen bromide fraction V was used as the main source of CB8 and CB10. CB7 was present in both cyanogen bromide fractions V and VI in approximately equal amounts. CB4 was purified from fraction VI. CB14, which contained no homoserine and corresponds to the carboxyl terminus of the protein, was the main peptide in cyanogen bromide fraction VII. Cyanogen bromide fraction VIII gave CB6 and CB9. CB9, purified by this route, was present in low yield. A better alternative was to isolate it from fraction VIII by high-voltage electrophoresis on paper.

Cyanogen bromide fractions IX and X were separated exclusively by paper electrophoresis. High-voltage electrophoresis (at pH 3.7 and 2.5 kV for 1 h) of fraction IX yielded three main peptides of which the most basic (mobility 20 cm to cathode) has the sequence Arg-Gly-Hse (CB3). Two other

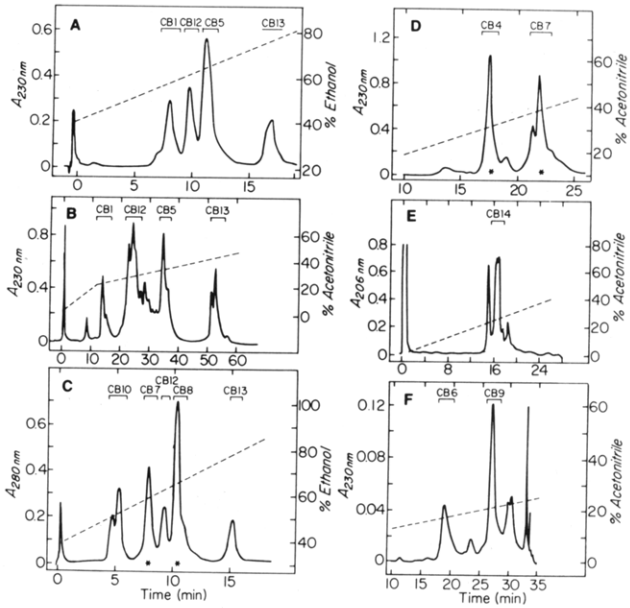

FIGURE 4: HPLC of cyanogen bromide fractions IV–VIII. Details of the chromatographic systems are described under Materials and Methods. Asterisks indicate radioactive peptides. (A) CB fraction IV, 1.7 mg, system 1, 1.5 mL/min; (B) CB fraction IV, 1.7 mg, system 2, 1 mL/min; (C) CB fraction V, 1.6 mg, system 1, 1.5 mL/min; (D) CB fraction VI, 850 μg, system 3, 1 mL/min; (E) CB fraction VII, 1.6 mg, system 2, 2 mL/min; (F) CB fraction VIII, 500 μg, system 4, 1 mL/min. Fractions were collected and pooled as indicated.

peptides (mobility 9 and 10.5 cm to cathode) had identical amino acid compositions and had the sequence Tyr-Gly-Gly-Hse (CB2). Analysis of cyanogen bromide fraction X by the same procedure also yielded CB2. Both CB fractions IX and X contained free homoserine. On one occasion, analysis of fraction X showed the presence of a peptide in poor yield (3%) with the composition Thr (0.73), Glu (1.00), and Hse (1.00). Later studies showed that a peptide of this composition would be expected if methionines-387 and -390 were both cleaved by cyanogen bromide. However, this cleavage reaction does not proceed in good yield at the Met-387–Thr-388 bond (Waxdal et al., 1968; Schroeder et al., 1969).

At the conclusion of the purification work, a total of 14 unique CB peptides were obtained in reasonable yield. Their

amino acid compositions are shown in Table I, and their sequence analysis is discussed in the next section. A total of four CB peptides (CB4, CB7, CB8, and CB11) contained [^{14}C]-carboxymethylcysteine residues.

Sequence Analysis. A complete summary of the sequence analyses of citrate synthase is presented in Figure 1. The proposed structure is best considered as three segments of continuous sequence from 1–46, 46–229, and 229–437. The proposed linkages between these three segments are still tentative since only single-residue overlaps were obtained; however, the order of the segments represents the only possible alignment. This is considered in more detail under Discussion.

(*A*) *Residues 1–46.* The sequence of the amino-terminal 36 residues was obtained from CB1, with tentative identifications of residues 28 and 31. Histidine-28 was confirmed by sequencing Tc3; however, threonine-31 was still tentative as was threonine-37. The threonine compositions of Tc3 (1.69 for 2) and CB1 (3.5 for 4) were consistent with the sequence assignments. The sequence of Tc3 placed methionine-41 between CB1 and CB2 and overlapped CB1 and CB2. Consistent with this sequence, free homoserine was detected among the peptides in cyanogen bromide fraction X. Methionine-45 was only identified tentatively by sequence; however, the isolation of CB2 and its composition require methionine at this residue.

(*B*) *Residues 46–229.* The overlap of CB3 and CB4 was established by the sequence of Tc4. The approximate location of CB4 in the citrate synthase molecule was also confirmed by sequencing HAα-AC1. This is a 20000-dalton polypeptide (residues 60–257) formed by acid treatment of HAα (residues 1–267). The overlap of CB4 and CB5 was confirmed by the sequence of both HAα-AC1 and Tc6.

For completion of the sequence of CB5 (residues 78–127), the peptide was digested with *S. aureus* V8 protease, and a total of four peptides were isolated. CB5-SP3 and CB5-SP4 were derived from the carboxyl-terminal region of CB5. The first two residues of CB5-SP3 overlapped the primary sequence of CB5. CB5-SP4 was the carboxyl-terminal peptide since it contained homoserine. The presence of tryptophan at residues 94, 109, and 114 was confirmed by BNPS-skatole degradation of Tc6 (residues 68–117) which liberated four main peptides. The composition and sequence of Tc6-W3 and Tc6-W4 confirm the correct ordering of CB5-SP3 and CB5-SP4.

CB5, CB6, and CB7 were placed in order by the overlapping sequence of Tc7. CB7 and CB8 were overlapped by the sequence of Tc9. Tc9 was a radioactive peptide, and Pth-[^{14}C]CMCys was released at cycle 11 (residue 175) and at cycle 20 (residue 184) of the degradation.

The sequence up to residue 229 was completed by the sequence of Tc11 which ordered CB8 and CB9. Again there is a minimal overlap of sequence at residues 228–229 due to the presence of the Met-Arg sequence in the proposed structure.

(*C*) *Residues 229–437.* CB10 was sequenced through 31 residues, revealing an aspartyl–proline bond at residues 257–258 which proved useful in completing the sequence in this region. Intially Tc12 (residues 230–313) was fragmented with cyanogen bromide, and the total mixture was analyzed to reveal only two sequences, Leu-230-Tyr-231-Leu-232- and Asn-267-Gly-268-Leu-269-, proving the ordering of CB10 and CB11. Tc12 was then succinylated to block its amino terminus and cleaved with acid at the 257–258 bond. This mixture was sequenced (Tc12-Ac1; Figure 1) without prior purification to complete the intervening sequence and establish the overlap of CB10 and CB11. CB11 was sequenced for a total of 61 residues; however, identification of Val-283 (turn 17) was missed. In addition the identifications of tryptophan at 284 and 306 were weak. For confirmation of these residues, CB11 was digested with BNPS-skatole, and the mixture was sequenced without prior purification. This revealed three sequences which were deduced as Asn-Gly-Leu-Ala-Gly-Pro-Leu-His-, Leu-Thr-Gln-Leu-Gln-Lys-Glu-Val-, and Asn-Thr-Leu-Asn-Ser-Gly-. The observation of these sequences is completely consistent with the tryptophan locations in this region of sequence. The fact that CB11 had the only asparaginyl–glycine bond in the protein enabled us to test the consistency of our proposed structure by hydroxylamine cleavage (Bornstein, 1970) of citrate synthase. This yielded two main polypeptides, Haα (27000–29000 daltons; residues 1–267) and HAβ (17000–19000 daltons; residues 268–437). The molecular mass of these fragments and the hydroxylamine cleavage pattern of the protein are in good agreement with the proposed structure of the enzyme (Figure 1). Both HAα and HAβ were isolated by gel filtration and sequenced. The amino-terminal sequence of HAβ confirmed Val-283, and this was reconfirmed by the sequence of Tc12-AC1.

The composition of CB11 revealed five arginine residues of which two (residues 302 and 313) were already placed in the primary sequence analysis. This means that several Tc peptides should be located in the carboxyl-terminal region. To order these peptides, we used fragments of citrate synthase (CS subt 37 and CS subt 10) generated by limited proteolysis with subtilisin (Bloxham et al., 1980). The apparent amino terminus of CS subt 10 was reported previously; however, it is now clear that the mixture of overlapping fragments was poorly interpreted. We repeated this analysis for a total of 19 residues but still failed to obtain positive residue identification at cycles 3 (residue 324), 5 (326), 8 (329), and 14 (335). The principal amino terminus of CS subt 10 (Val-Leu) overlapped the established sequence of CB11 and Tc13. Additional confidence in this overlap was established by isolation of Tc13′ from a tryptic digest of CS subt 37 which differed from Tc13 by the loss of the carboxyl-terminal residues Val-Leu-Arg, consistent with the site of subtilisin cleavage and indicating an arginine as the missing residue 324. The overlap of the sequence of CS subt 10 with Tc15 and Tc16 was clearly established which left just two residues to account for. The formation of Tc15 requires arginine at residue 329. Therefore, Tc14 is the only tryptic peptide with a sequence which can be accommodated between residues 325 and 329, and threonine must be located at residue 326. The location of cysteine at residue 332 agrees with the fact that CB11 contains a [^{14}C]CM group.

Cyanogen bromide digestion of CS subt 10 yielded four peptides by gel filtration, of which three (CB12, CB13, and CB14) had been isolated from the intact protein and the fourth contained the amino terminus of CS subt 10 (residues 322–346). Therefore, fragments CB12, CB13, and CB14 can be unambiguously placed in the carboxyl-terminal region of the protein. Since CB14 is the only cyanogen bromide peptide lacking homoserine, it must contain the actual carboxyl terminus. Acid cleavage of succinylated CS subt 10, followed by dialysis to remove small peptides, yielded a polypeptide (CS subt 10-AC1) of approximately 10000 daltons (residues 345–437). Its amino terminus overlapped CB12, and hence the order of the final CB peptides could be deduced as CB12-CB13-CB14. This was confirmed by the overlap of the Tc16 and CB12 sequences.

Table I: Amino Acid Composition of Cyanogen Bromide Fragments[a]

	CB1	CB2	CB3	CB4	CB5	CB6	CB7	CB8	CB9	CB10	CB11	CB12	CB13	CB14
residues	1–40	42–45	46–48	49–77	78–127	128–138	139–176	177–216	217–228	229–266	267–346	347–390	391–425	426–437
Sephadex fraction	IV	X	X	VI	IV	VIII	V + VI	V	VIII	V	III	IV + V	IV + V	VII
subsequent purification	HPLC	HVE[b]	HVE	HPLC	HPLC	HVE or HPLC	HPLC	HPLC	HPLC	HPLC	SPS[c]	HPLC	HPLC	HPLC
CMCys				0.5 (1)			1.0 (1)	0.91 (1)			0.71 (1)			
Asx	4.41 (5)			1.96 (2)	1.14 (0)	3.09 (3)	3.15 (3)	5.15 (6)	1.06 (1)	2.98 (3)	8.76 (9)	4.14 (4)	1.23 (1)	2.01 (2)
Thr	3.5 (4)			1.03 (1)	3.2 (3)	0.8 (1)	1.87 (2)	1.4 (1)	1.57 (2)	2.09 (2)	3.82 (4)	1.29 (1)	1.23 (1)	0.95 (1)
Ser	2.4 (2)			2.13 (2)	3.34 (3)	1.49 (0)	3.05 (4)	3.6 (4)		5.7 (6)	2.05 (2)	1.19 (1)	3.08 (3)	1.97 (2)
Hse	0.68 (1)	1.09 (1)	1.0 (1)	1.05 (1)	0.7 (1)	0.55 (1)	0.65 (1)	0.4 (1)	0.4 (1)	0.94 (1)	0.68 (1)	0.9 (2)	0.4 (1)	
Glx	5.16 (5)			3.67 (4)	7.88 (8)		4.38 (5)	1.19 (1)	2.21 (2)	1.38 (1)	8.4 (8)	5.31 (5)	2.31 (2)	
Pro	1.09 (1)			1.05 (2)	5.22 (5)	1.49 (2)		1.4 (1)		0.74 (1)	4.69 (5)	2.77 (3)	1.94 (2)	
Gly	2.01 (2)	2.0 (2)	1.0 (1)	3.55 (3)	4.5 (4)		1.37 (1)	2.47 (2)	1.3 (1)	3.32 (3)	6.89 (7)	3.21 (3)	3.08 (3)	1.27 (1)
Ala	3.14 (3)			0.3 (0)	4.43 (4)		4.98 (6)	4.12 (4)	1.12 (1)	4.7 (5)	4.0 (4)	2.82 (3)	2.66 (3)	
Val	2.23 (3)			1.8 (2)	4.01 (4)			1.35 (1)		2.34 (2)	7.05 (7)	4.5 (5)	2.96 (3)	0.86 (1)
Ile	3.11 (4)			1.62 (2)	1.48 (1)		2.66 (3)	3.56 (4)		0.82 (1)	0.98 (1)	0.99 (1)	1.0 (1)	1.08 (1)
Leu	3.02 (3)			1.7 (2)	7.73 (7)	1.86 (2)	3.5 (3)	4.66 (4)	2.2 (2)	4.92 (5)	10.89 (11)	5.7 (6)	5.74 (6)	2.0 (2)
Tyr	0.35 (0)	0.57 (1)		1.65 (2)			1.72 (3)	1.86 (2)	0.98 (1)	1.81 (2)	3.15 (3)	2.43 (3)	1.54 (2)	
Phe	1.02 (1)			0.66 (1)	0.9 (1)	0.7 (1)	1.22 (1)	1.03 (1)	1.05 (1)	1.11 (1)	1.24 (1)	1.19 (1)	1.77 (2)	
His	1.31 (1)				1.13 (1)	0.7 (1)	1.28 (1)	1.42 (1)		4.18 (4)	3.04 (4)	1.09 (1)		
Lys	2.75 (3)			1.93 (2)	3.65 (4)		1.01 (1)	2.43 (3)			5.15 (5)	3.76 (4)[d]	0.93 (1)	1.66 (2)
Arg	2.10 (2)		0.56 (1)	2.2 (2)	1.26 (1)		2.41 (2)	2.39 (2)		1.26 (1)	5.22 (5)		3.25 (3)	
Trp					ND[e] (3)		ND (1)	ND (1)			ND (2)	ND (1)	ND (1)	
no. of residues	40	4	3	29	50	11	38	40	12	38	80	44	35	12
% yield	45	36	40	53	75	49	30	80	25	75	50	65	55	60

[a] Residues per molecule by amino acid analysis after 18-h hydrolysis or (in parentheses) from the sequence (Figure 5). [b] HVE, high-voltage electrophoresis on paper at pH 3.7 and 2.5 kV for 1 h. [c] SPS, SP Sephadex C-25 chromatography. [d] 0.79 mol of trimethyllysine was identified per mol of peptide CB12. [e] Not determined.

CB12 was sequenced for 36 cycles, but identifications were lacking at cycles 22, 25, and 31. Subsequently CB12 was digested with trypsin which yielded four peptides. The sequence analysis of CB12-T4 for 23 residues led to the identification of trimethyllysine (residue 368) at the second cycle of degradation. This was established by comparison of the mobility of the Pth derivative with that of authentic Pth-trimethyllysine and with the derivative of residue 115 from human brain calmodulin which is also trimethyllysine (Schreiber et al., 1981). Amino acid analysis of CB12 and CB12-T4 under conditions which resolve lysine and trimethyllysine confirmed the prescence of one residue of trimethyllysine.

The terminal sequence of CB12-T4 unexpectedly indicated homoserine at cycle 21 prior to the termination of the sequence. To check this, we digested CB12-T4 with chymotrypsin, and the sequence of CB12-T4-C1 was established as Gly-Hse-Thr-Glu-(Hse). Clearly cyanogen bromide cleavage at Met-387 was occurring in very poor yield due to the presence of Thr-388. This was confirmed by cyanogen bromide digestion of Tc16 and sequencing the total mixture. The major sequences were Glu-Phe-, Phe-Lys-, and Asn-Tyr-. Thr-Glu-(Tc16-CBα) was only observed as a minor sequence (5–10% yield). The overlap between CB12 and CB13 was established by the sequence of Tc16-W1 which also served to give positive identifications of Met-387 and Met-390 as well as confirming the presence of tryptophan at residue 371.

The sequence was completed by analyzing CB14, which gave poor identification of the last two residues and Lys-432. The latter residue was positively confirmed in the sequence of Tc17 which overlapped both CB13 and CB14. Finally the sequence of a tryptic peptide CB14-T2 allowed confident identification of the carboxyl terminus as Ser-436-Lys-437.

An earlier paper (Bloxham et al., 1981b) demonstrated that this 437-residue amino acid sequence is in accord with the amino acid composition of acid hydrolysates of the protein.

Discussion

Citrate synthase contains 15 methionine and 19 arginine residues (Bloxham et al., 1981b), and theoretically, cleavage at these residues should yield a maximum of 16 and 20 fragments, respectively. Either of these mixtures would be rather complex to fractionate into homogeneous peptides by traditional techniques. We first explored a strategy based upon specific limited cleavage by proteolysis and by chemical means at asparaginyl–glycine bonds or aspartyl–proline bonds to produce smaller defined fragments prior to using more general cleavage procedures (Walsh et al., 1978, 1981). The successful development of HPLC for peptide purification [e.g., see Gerber et al. (1979), Hughes et al. (1979), and Mahoney & Hermodson (1980)] has redirected this strategy since proper use of this technique enables rapid and sensitive peptide isolation from rather complex mixtures. This is clearly shown by the purification of the CB peptides (11 of 14), the final resolution of tryptic peptides, and the purification of subdigest peptides from citrate synthase. Many of these peptides were not separable by ion-exchange chromatography (i.e., CB8 and CB10; Tc9 and Tc17). The HPLC technology is particularly useful for peptides containing 10–50 residues, the typical size range of the peptides isolated from citrate synthase. Nevertheless, it is worthwhile noting that the products of specific cleavage reactions were invaluable in ordering some peptides in the sequence and in confirming the proposed final structure.

Every determination of an amino acid sequence has areas of relative uncertainty, and the present case is no exception. The sequence of citrate synthase has been unambiguously established in three continuous segments, and the deduced

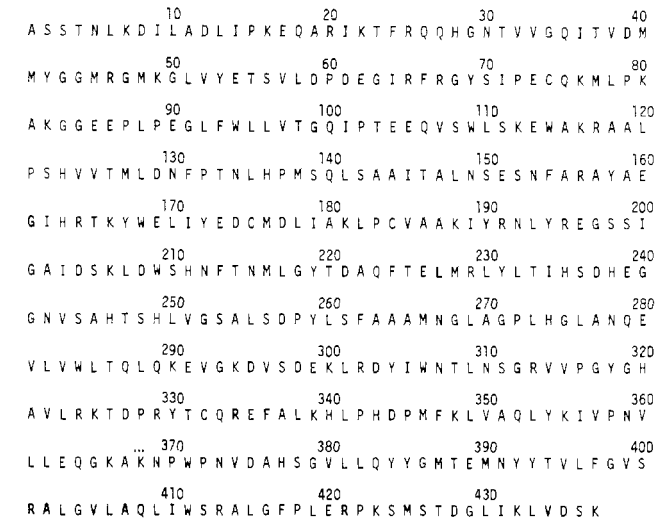

FIGURE 5: Amino acid sequence of porcine heart citrate synthase. The one-letter code abbreviations are given in Figure 1.

structure is shown in Figure 5. Fewer peptides were isolated from the cyanogen bromide and tryptic digests than are predicted. However, the differences are explained by cleavage at Met-40–Met-41 to yield free homoserine, failure of cyanogen bromide to cleave Met-387–Thr-388, and failure of trypsin to cleave Arg-302–Asp-303 [cf. Konigsberg & Steinman (1977)] and Arg-421–Pro-422 [cf. Tsugita et al. (1960)]. There is no obvious reason why Arg-401–Ala-402 is virtually resistant to trypsin. Identification of both these residues in the Edman degradation of CB13 was positive. Gel filtration experiments showed that Tc16 (residues 335–413), which contains the resistant bond, eluted with a much higher apparent molecular mass than Tc12 (residues 230–313), which is in fact slightly larger. It is clear that the region of Tc16 must form a high molecular weight aggregate which might prevent tryptic attack.

The ordering of Cb peptides was achieved by overlapping sequences in the appropriate tryptic peptides. Generally a substantial sequence overlap was obtained; however, at residues 45–46 and 228–229, there was a minimal overlap of one residue due to the two Met-Arg sequences. The three segments of sequence (residues 1–46, 46–229, 229–437) are placed in their correct order since the first segment (residues 1–46) contains the amino terminus of the intact protein and the last segment (residues 229–437) contains the carboxyl terminus of the protein. In the regions of minimal overlap, the only alternative sequences are as follows: -Met-$(X)_n$-Arg-; -Met-$(Arg)_n$-Arg-; -Met-Arg-$(X)_n$-Met-Arg-; -Met-$(Met)_n$-Arg-. The first three possibilities are eliminated by the failure to identify, despite exhaustive analysis, any unaccounted peptide sequence in cyanogen bromide fragments. The fourth solution is still feasible; however, the electron density of the X-ray crystal structure in the region of both minimal overlaps is in complete accord with the proposed sequence (S. J. Remington, G. Wiegand, and R. Huber, unpublished experiments).

In an ideal sequence determination, all residues would be positively identified from the sequence of at least two different peptides. For citrate synthase, this criterion was fulfilled for 328 residues whereas 109 residues were identified from single peptide sequences. Many of these residues were confirmed by repeated analysis; however, in the regions 37–41 and 260–266, the sequence relies on a single analysis. At the current point, only Thr-31, Thr-37, and Met-45 have weak identifications by sequence analysis. Methionine-45 is required

for the formation of CB3, and both threonine residues are in accord with the appropriate amino acid analyses.

The presence of trimethyllysine-368 in citrate synthase was totally unexpected. Although this derivative is found in many animal proteins including histones (Hempel et al., 1968), myosin (Hardy et al., 1970), ribosomal proteins (Chang et al., 1976), and calmodulin (Vanaman et al., 1977), it has never been identified in an intramitochondrial protein. Indeed, although trimethyllysine is found in prokaryotic and yeast cytochrome *c*, the corresponding residue in animal cytochrome *c* is always lysine (DeLange et al., 1970). In rat, the heart contains the highest concentration of trimethyllysine of any of the tissues (Paik & Kim, 1980), and presumably protein methylase (Paik & Kim, 1970, 1980) is active in this tissue. However, since citrate synthase is coded by nuclear DNA (Van Heyningen et al., 1973) and the protein methylases are present in the nucleus (Paik & Kim, 1970) and cytosol (Kim et al., 1978), modification must occur before citrate synthase accumulation in the mitochondria. It would be especially interesting to know if methylation influences the transport into the mitochondrion, the catalytic activity, or both. Preliminary evidence from the X-ray crystal structure suggests that trimethyllysine-368 is in the region of the active site (S. J. Remington, G. Wiegand, and R. Huber, unpublished experiments).

The amino aid sequence determined in this work has been used as a basis to refine the X-ray crystal structure of the enzyme (S. J. Remington, G. Wiegand, and R. Huber, unpublished experiments), and this has allowed identification of a number of features of the enzyme structure indicating potentially important catalytic groups. These two studies together will form the molecular basis for detailed investigations of the mechanism of carbon–carbon bond formation by the enzyme and the regulation of its catalytic activity.

Acknowledgments

We thank Richard R. Granberg for performing the amino acid analyses and both Roger D. Wade and Lowell H. Ericsson for their assistance in the Edman degradation procedure. We are indebted to Dr. Hans Neurath for his interest and helpful discussions during the course of this work.

Supplementary Material Available

Descriptions of the purification and amino acid compositions of peptides used in parts of this work in 3 supplemental tables and 11 figures (16 pages). Ordering information is given on any current masthead page.

References

Atassi, M. Z., & Habeeb, A. F. S. A. (1972) *Methods Enzymol. 25*, 546.

Bloxham, D. P., Ericsson, L. H., Titani, K., Walsh, K. A., & Neurath, H. (1980) *Biochemistry 19*, 3979.

Bloxham, D. P., Parmelee, D. C., Kumar, S., Wade, R. D., Ericsson, L. H., Walsh, K. A., Neurath, H., & Titani, K. (1981a) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 40*, 1677.

Bloxham, D. P., Parmelee, D. C., Kumar, S., Wade, R. D., Ericsson, L. H., Neurath, H., Walsh, K. A., & Titani, K. (1981b) *Proc. Natl. Acad. Sci. U.S.A. 78*, 5381.

Bornstein, P. (1970) *Biochemistry 9*, 2408.

Brauer, A. W., Margolies, M. N., & Haber, E. (1975) *Biochemistry 14*, 3029.

Chang, F. N., Navickas, I. J., Chang, C. N., & Dancis, B. M. (1976) *Arch. Biochem. Biophys. 172*, 627.

DeLange, R. J., Glazer, A. N., & Smith, E. L. (1970) *J. Biol. Chem. 245*, 3325.

Edman, P., & Begg, G. (1967) *Eur. J. Biochem. 1*, 80.

Ericsson, L. H., Wade, R. D., Gagnon, J., McDonald, R. M., Granberg, R. R., & Walsh, K. A. (1977) in *Solid Phase Methods in Protein Sequence Analysis* (Previero, A., & Coletti-Previero, M. A., Eds.) p 137, Elsevier/North-Holland, Amsterdam.

Gerber, G. E., Anderegg, R. J., Herlihy, W. C., Gray, C. P., Biemann, K., & Khorana, H. G. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 227.

Hardy, M. F., Harris, C. I., Perry, S. V., & Stone, D. (1970) *Biochem. J. 120*, 653.

Hempel, K., Lange, H. W., & Birkofer, L. (1968) *Naturwissenschaften 55*, 37.

Hermann, J., Titani, K., Ericsson, L. H., Wade, R. D., Neurath, H., & Walsh, K. A. (1978) *Biochemistry 17*, 5672.

Hermodson, M. A., Ericsson, L. H., Titani, K., Neurath, H., & Walsh, K. A. (1972) *Biochemistry 11*, 4493.

Hughes, G. J., Winterhalter, K. H., & Wilson, K. J. (1979) *FEBS Lett. 108*, 81.

Kim, S., Nochumson, S., Chin, W., & Paik, W. K. (1978) *Anal. Biochem. 84*, 415.

Koide, A., Titani, K., Ericsson, L. H., Kumar, S., Neurath, H., & Walsh, K. A. (1978) *Biochemistry 17*, 5657.

Konigsberg, W. H., & Steinman, H. M. (1977) *Proteins (3rd Ed.) 3*, 1–178.

Laemmli, U. K. (1970) *Nature (London) 227*, 680.

Lynen, F., & Reichert, S. (1951) *Angew. Chem. 63*, 47.

Lynen, F., Reichert, S., & Rueff, L. (1951) *Justus Liebigs Ann. Chem. 574*, 1.

Mahoney, W. C., & Hermodson, M. A. (1980) *J. Biol. Chem. 255*, 11199.

Ochoa, S. (1954) *Adv. Enzymol. Relat. Areas Mol. Biol. 15*, 183.

Ochoa, S., Stern, J. R., & Schneider, M. C. (1951) *J. Biol. Chem. 193*, 691.

Paik, W. K., & Kim, S. (1970) *J. Biol. Chem. 245*, 6010.

Paik, W. K., & Kim, S. (1980) in *Protein Methylation*, p 21, Wiley, New York.

Piszkiewicz, D., Landon, M., & Smith, E. L. (1970) *Biochem. Biophys. Res. Commun. 40*, 1173.

Schreiber, W. E., Sasagawa, T., Titani, K., Wade, R. D., Malencik, D., & Fischer, E. H. (1981) *Biochemistry 20*, 5239.

Schroeder, W. A., Shelton, J. B., & Shelton, J. R. (1969) *Arch. Biochem. Biophys. 130*, 551.

Singh, M., Brooks, G. C., & Srere, P. A. (1970) *J. Biol. Chem. 245*, 4636.

Spector, L. B. (1972) *Enzymes, 3rd Ed. 7*, 357–389.

Srere, P. A. (1969) *Methods Enzymol. 13*, 3.

Srere, P. A. (1975) *Adv. Enzymol. Relat. Areas Mol. Biol. 43*, 57.

Stern, J. R., Shapiro, B., & Ochoa, S. (1950) *Nature (London) 166*, 403.

Stern, J. R., Shapiro, B., Stadtman, E. R., & Ochoa, S. (1951) *J. Biol. Chem. 193*, 703.

Tarr, G. E., Beecher, J. F., Bell, M., & McKean, D. J. (1978) *Anal. Biochem. 84*, 622.

Tsugita, A., Gish, D. T., Young, J., Fraenkel-Conrat, H., Knight, C. A., & Stanley, W. M. (1960) *Proc. Natl. Acad. Sci. U.S.A. 46*, 1463.

Vanaman, T. C., Sharief, F., & Watterson, D. M. (1977) in *Calcium Binding Proteins and Calcium Function* (Wasserman, R. H., Corrandino, R. A., Carafoli, E., Kretsinger, R. H., MacLennan, D. H., & Siegel, F. L., Eds.) pp

107-116, North-Holland, New York.
Van Eldik, L. J., Grossman, A. R., Iverson, D. B., & Watterson, D. M. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1912.
Van Heyningen, V., Craig, I., & Bodmer, W. (1973) *Nature (London) 242*, 509.
Walsh, K. A., Ericsson, L. H., & Titani, K. (1978) in *Versatility of Proteins* (Li, C. H., Ed.) pp 39-58, Academic Press, New York.
Walsh, K. A., Ericsson, L. H., Parmelee, D. C., & Titani, K. (1981) *Annu. Rev. Biochem. 50*, 261.
Waxdal, M. J., Konigsberg, W. H., Henley, W. L., & Edelman, G. M. (1968) *Biochemistry 7*, 1959.
Weitzman, P. D. J., & Danson, M. J. (1976) *Curr. Top. Cell. Regul. 10*, 161.
Wiegand, G., Kukla, D., Scholze, H., Jones, T. A., & Huber, R. (1979) *Eur. J. Biochem. 93*, 41.

Analysis of an Allosteric Binding Site: The Nucleoside Inhibitor Site of Phosphorylase *a*†

Stephen Sprang, Robert Fletterick,* Michael Stern, Daniel Yang,‡ Neil Madsen,§ and Julian Sturtevant‖

ABSTRACT: Glycogen phosphorylase is inhibited by a family of related compounds [purines, purine nucleosides, nucleotides, and certain heterocyclic compounds, e.g., flavin mononucleotide (FMN)] which bind to an allosteric site located at the surface of the enzyme 10 Å from the catalytic cleft at which glucose and glucose 1-phosphate are bound. The interaction of several such inhibitors, adenine, caffeine, adenosine, inosine, ATP, and FMN, with rabbit muscle phosphorylase *a* in the glucose-inhibited form has been examined by X-ray crystallographic (difference Fourier) analysis at 3.0- and 2.5-Å resolution. The dissociation constant (K_d) for all of these ligands was determined by kinetic analysis and, for FMN, by fluorometry. The $\Delta S°$ and ΔH for association of FMN to phosphorylase were derived from analysis of calorimetric data. We have synthesized the structural and thermodynamic data in order to arrive at a description of the energetics of the binding interaction and its specificity. The inhibitors associate with phosphorylase by forming an intercalative complex in which the heterocyclic ring system is stacked between the aromatic side chains of F285 and Y612. When they do so, the inhibitor stabilizes the same conformation of residues 282-286 which binds α-D-glucose, an inhibitor which demonstrates synergism with the purine ligands. No other significant hydrogen-bonded contacts are made with the enzyme, and any polar or charged groups of the ligand (ribose, ribose phosphate, or ribityl phosphate) are solvated at the protein surface. There does not appear to be a single particularly favored orientation of the heterocyclic ring dipole within its binding pocket nor is there a favored orientation for the polar moiety of the ligand at the surface of the enzyme. Association free energies range from 2.0 kcal (ATP) to 7.0 kcal (FMN), and the complete quenching of FMN fluorescence on binding suggests that the stacking interaction is quite strong. Calorimetric analysis of FMN binding reveals that both ΔH and $\Delta S°$ of association are significant. An unusual feature of the interaction is the strong temperature dependence of ΔC_p. Our analysis of the complex demonstrates that ΔG_d is not a simple function of the loss of accessible surface for protein and ligand upon binding. On the other hand, loss of accessible surface of the heterocyclic ring system alone correlates directly with ΔG_d. We conclude that the association energy derives solely from attractive dispersion forces in which both enthalpic and entropic contributions are significant. Accordingly, change in accessible surface on binding reflects both contributions to the free energy of binding.

Crystallographic analysis has revealed many examples of the intricate network of nonbonded interactions which characterize protein-ligand complexes. However, the balance of forces which confer stability or specificity to these interactions remains poorly understood. Such complexes have rather low free energies of association (3-10 kcal/mol) because they arise from the small differences among many individual entropic and enthalpic contributions of much greater magnitude. Structural analysis has allowed direct visualization of the role of electrostatics and complementarity. On the other hand, the static structure does not suggest the entropic contributions, viz., changes in translational and rotational degrees of freedom in protein, ligand, and solvent, that result from binding. Recent attempts have been made to evaluate entropic (hydrophobic) contributions from structural data by computing the solvent-accessible surface area (ΔA) lost by protein and ligand as a result of binding (Janin & Chothia, 1978). However, this prescription is based on thermodynamic analysis of phase equilibria in simple systems (Nozaki & Tanford, 1971) which bear little resemblance to the complex environment within a protein-ligand binding pocket (Hvidt, 1978). The real contributions of electrostatics or buried surface to stability can be more accurately evaluated if ΔH and ΔS can be directly determined for the systems of interest (Sturtevant, 1977). In this report, we discuss the results of a combined crystallographic and thermodynamic approach to describe the ener-

† From the Department of Biochemistry and Biophysics, School of Medicine, University of California, San Francisco, California 94143. *Received September 22, 1981; revised manuscript received January 14, 1982.* This work has been supported by grants from the Medical Research Council of Canada (to N.M.), the University of Alberta Computer Center, the National Institutes of Health (Grant AM 26081 to R.F. and Grant GM 04725 to J.S.), NIH Postdoctoral Traineeship (Grant AM 06293 to S.S.), the National Science Foundation (Grant PCM79-06000 to R.F.), and the Kroc Foundation.

‡ Present address: Department of Crystallography, University of Pittsburgh, Pittsburgh, PA.
§ Present address: Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2H7.
‖ Present address: Department of Chemistry, Yale University, New Haven, CT 06510.

0006-2960/82/0421-2036$01.25/0