Sequence of the Carboxyl-Terminal 492 Residues of Rabbit Muscle Glycogen Phosphorylase Including the Pyridoxal 5′-Phosphate Binding Site†

Koiti Titani,‡ Atsushi Koide,§ Lowell H. Ericsson, Santosh Kumar, Jacques Hermann, Roger D. Wade, Kenneth A. Walsh,* Hans Neurath, and Edmond H. Fischer

Appendix I: Crystallographic Analysis of Phosphorylase *a* at 2.5 Å Resolution, a Comment on the Chemical Sequence‖

Stephen Sprang* and R. J. Fletterick

Appendix II: Assignment of the Amino Acid Sequence to the Crystal Structure of Glycogen Phosphorylase *b*#

J. A. Jenkins,* L. N. Johnson, and K. S. Wilson

ABSTRACT: This communication presents the strategy and experimental details which establish the amino acid sequence of the carboxyl-terminal 492 residues (residues 350 through 841) of rabbit muscle glycogen phosphorylase (EC 2.4.1.1). The heavy segment (H_s), derived from the native enzyme by limited proteolysis with subtilisin, was cleaved with cyanogen bromide to yield 15 fragments. The amino acid sequences of 12 of these are described herein. The sequence of 3 other fragments (CB17C, CB18, and CB15) is described in accompanying reports by Koide, A., et al., and Hermann, J., et al. ((1978) *Biochemistry 17* (first and second papers, respectively, in a series in this issue)). These 15 fragments were aligned by analysis of three others generated by cleavage of the heavy segment H_s at asparaginylglycine bonds with hydroxylamine and of four more generated by acid cleavage of aspartylproline bonds. Lysine-679 was identified as the binding site of the essential cofactor pyridoxal 5′-phosphate. These data, together with those reported in the accompanying papers (vide supra), establish the complete sequence of the 841 amino acid residues in glycogen phosphorylase. They provide a chemical basis on which the relationship between structure and function of the enzyme can be examined.

Rabbit muscle glycogen phosphorylase is a homooligomer in which each subunit contains a single chain of 841 amino acid residues. In order to facilitate its sequence analysis, the protein was cleaved by subtilisin into a light and a heavy segment (L_s and H_s) by Koide et al. (1978), who also described the sequence of the amino-terminal segment, L_s, and that of a fragment (CB17) overlapping the two segments. The sequences of two large cyanogen bromide fragments (CB18 and CB15) comprising 78 and 163 residues, respectively, are reported by Hermann et al. (1978). In the present communication these 2 fragments are aligned with the sequences of the amino-terminal segment L_s of Koide et al. (1978) and of 12 additional cyanogen bromide fragments to complete the proof of the sequences of both segment H_s (residues 265–841) and the entire molecule, as summarized in a preliminary communication (Titani et al., 1977).

Materials and Methods

Segment H_s was prepared from reduced, carboxymethylated rabbit muscle glycogen phosphorylase *b* as described by Koide et al. (1978).

Methods of cleavage of peptide bonds at methionine, arginine, tryptophan, and aspartylproline residues were described previously, as well as methods of enzymatic digestion with trypsin, chymotrypsin, thermolysin, carboxypeptidase, staphylococcal protease and pepsin (Koide et al., 1978). In certain cases cleavage at methionyl residues by thermolysin was deliberately prevented by prior oxidation to methionine sulfone with performic acid (Hirs, 1967). Asparaginylglycine bonds were cleaved by the following modification of the method of Bornstein (1969). The protein (1–10 mg/mL) was dissolved in 2 M hydroxylamine hydrochloride containing 6 M guanidine hydrochloride and maintained at pH 9.0 (45 °C) by the addition of 4.5 M LiOH in a pH-stat for 6 h. The reaction mixture was then adjusted to pH 2–3 with formic acid and desalted either by dialysis against water or by gel filtration in 9% formic acid.

Large fragments were separated by gel filtration, acid precipitation, or ion-exchange chromatography on SP-Sephadex as indicated. Smaller peptides were separated by combinations of gel filtration, ion-exchange chromatography on Bio-Rad AG1-X2 (Wikler et al., 1970) and paper electrophoresis (40 V/cm, 1–1.5 h) at pH 3.7 or 6.5 (Ryle et al., 1955; Bennett, 1967). Arginine-containing peptides were detected either on a strip of Whatman 3MM filter paper to which aliquots (1–2 nmol) of chromatographic fractions had been applied or on a guide strip from a paper electropherogram using phen-

† From the Department of Biochemistry, University of Washington, Seattle, Washington 98195. *Received July 6, 1978.* This work has been supported by research grants from the National Institutes of Health (GM-15731 and AM-7902).

‡ Investigator of the Howard Hughes Medical Institute.

§ Present address: Eisai Research Laboratories, Eisai Co., Ltd., Koishikawa 4, Bunkyo-ku, Tokyo, Japan.

‖ Present address: Institut Pasteur, 28 Rue du Docteur Roux, Paris 75015, France.

From the MRC Group on Protein Structure and Function and the Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2H7.

From the Laboratory of Molecular Biophysics, Department of Zoology, Oxford OX1 3PS, United Kingdom.

0006-2960/78/0417-5680$01.00/0

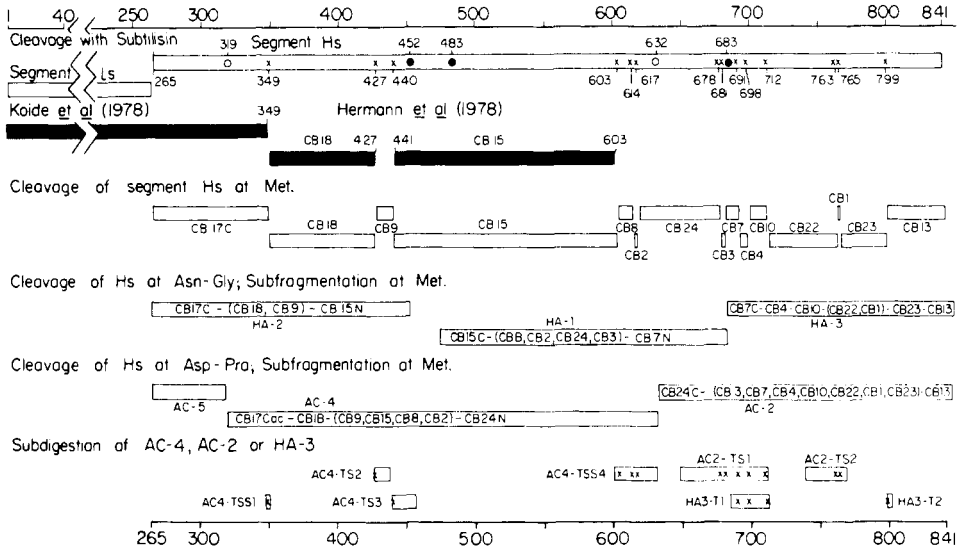

FIGURE 1: Diagrammatic summary of the strategy of analysis of the sequence of segment H_s (residues 265–841) derived from rabbit muscle phosphorylase *b* by subtilisin cleavage. Koide et al. (1978) have described the proof of sequence of residues 1–349 which includes segment L_s (residues 17–264) and fragment CB17 (residues 242–349). Hermann et al. (1978) have described the proof of sequence of fragments CB18 and CB15. Loci of cleavage (and residue numbers) are indicated in the upper bar by the following symbols: (O) Asp-Pro; (X) methionine; (●) Asn-Gly. The lengths of the fragments isolated are proportional to the lengths of the bars. The fragments obtained by cleavage with hydroxylamine (HA-) at Asn-Gly bonds or with acid (AC-) at Asp-Pro bonds are represented by bars within which are listed constituent subfragments generated by treatment of the parent fragments with CNBr. In these cases parentheses enclose subfragments in unknown order, whereas hyphens indicate proven order. A lettered suffix on a fragment number (e.g., CB17*N*) indicates that the fragment is either the amino-terminal portion (N) or the carboxyl-terminal portion (C) of a parent fragment (in this case, of CB17).

anthrenequinone (Eastman) according to the method of Yamada & Itano (1966). Tryptophan-containing peptides were detected on paper strips by the method of Smith (1960).

Automated Edman degradations were performed as described by Koide et al. (1978) for the liquid phase mode in a Beckman Sequencer or by Hermann et al. (1978) for the solid phase mode in a Sequemat. In specified cases small peptides were retained in the Sequencer by inclusion of 3 mg of the inert carrier Polybrene (Tarr et al., 1978). Other small peptides were degraded manually and residues identified either by the subtractive method of Konigsberg & Hill (1962) or by direct gas–liquid chromatographic analysis of the phenylthiohydantoins (Titani et al., 1975).

Results

The general strategy for the sequence determination of the carboxyl-terminal two-thirds (residues 265 through 841) of rabbit muscle phosphorylase subunit is summarized in Figure 1. Cleavage of segment H_s with cyanogen bromide yielded 15 fragments. The sequences of three of these (CB17C, CB18, and CB15) are described in accompanying reports (Koide et al., 1978; Hermann et al., 1978); the sequences of the 12 remaining fragments were determined as described below. In order to align these 15 fragments, segment H_s was cleaved at three asparaginylglycine and two aspartylproline bonds and 6 major fragments (HA-1, HA-2, HA-3, AC-2, AC-4, and AC-5) were isolated and analyzed. The remaining alignments were established by analysis of methionine-containing peptides derived by cleavage at arginyl or glutamyl residues.

Separation of Fragments of Segment H_s Generated by Cleavage with Cyanogen Bromide. The heavy segment H_s (800 mg; ca. 12 μmol) was cleaved with cyanogen bromide as described by Koide et al. (1978). The products were dissolved in 3 mL of 88% formic acid, diluted to 30 mL with water, and applied to a column of Sephadex G-50. The elution profile is shown in Figure 2A. Seventy-five percent (by weight) was recovered in four major fractions, i.e., fractions I (220 mg), II (235 mg), III (95 mg), and IV (50 mg). Fraction V was not weighed because it was slightly oily. Of these five fractions, only fractions I and IV were seemingly homogeneous. Nevertheless, they were further purified by rechromatography on the same column. Fraction I contained fragment CB15 (ca. 18 000 daltons) which had been previously isolated from the intact molecule by Saari & Fischer (1973); its sequence is reported by Hermann et al. (1978). Fraction IV contained a fragment (designated CB23) which had not been isolated before.

$NaDodSO_4$ gel electrophoresis of fraction II indicated the presence of three fragments of ca. 7000, 9000, and 10 000 daltons which were separated on a SP-Sephadex C-25 column in the presence of 7 M urea (Figure 2B) yielding fractions II-1 (25 mg), II-2 (40 mg), and II-3 (60 mg). Fraction II-1 contained another new fragment (CB24) which had been overlooked in earlier experiments due to its low absorbance at 280 nm. Fractions II-2 and II-3 were identified as fragment CB17C [the carboxyl-terminal portion of CB17 previously placed in sequence by Koide et al. (1978)] and CB18, the sequence of which is reported by Hermann et al. (1978). Fraction III, containing two fragments of ca. 5000 and 6000 daltons, was also further purified on a SP-Sephadex C-25 column (Figure 2C), yielding fractions III-1 (45 mg) and III-2 (40 mg). A third new fragment (CB22) was found in fraction III-1, whereas fraction III-2 was identical with fragment CB13 previously isolated by Saari & Fischer (1973).

It was observed that fractions 175 through 190 in Figure 2A contained an acid-insoluble precipitate which appeared during or following chromatography. It was collected by centrifugation of fraction V, following concentration to 50 mL by rotary evaporation, and washed three times with 5-mL aliquots of 9% formic acid. The material obtained (15 mg) was found to be identical in composition with fragment CB10 of Saari & Fischer (1973). The supernatant solution and washings were combined, lyophilized, and further separated on a Sephadex G-25 column (Figure 2D). Fraction V-1 contained pure fragment CB9. Fractions V-2, V-3, and V-4 were further purified on a column of AG 1-X2 using pyridine–acetate buffers

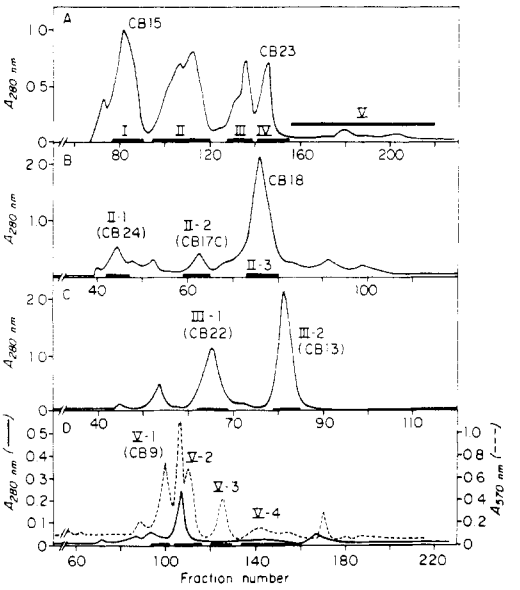

FIGURE 2: (A) Fragments generated by cleavage of segment H_s (800 mg) with cyanogen bromide were separated on a column (5.0 × 145 cm) of Sephadex G-50 superfine in 9% formic acid. Fractions of 15 mL were collected at 90 mL/h and pooled as indicated by the horizontal bars. (B) Further purification of fraction II (from A) on a column (1.5 × 60 cm) of SP-Sephadex C-25 at room temperature with a linear gradient from 0.1 M sodium formate (pH 2.9) containing 7 M urea to 0.75 M sodium formate (pH 3.9) containing 7 M urea (each 300 mL). Fractions of 5 mL were collected at 25 mL/h. Fragments were subsequently separated from urea and salts on a column (2.5 × 40 cm) of Sephadex G-25 Fine in 9% formic acid and lyophilized. (C) Purification of fraction III (from A) as in B, except that the gradient proceeded only to 0.5 M sodium formate (pH 3.5) in 7 M urea. (D) Separation of fraction V from A after removal of an acid-insoluble fragment (see text). A column (2.5 × 200 cm) of Sephadex G-25 superfine was used with 9% formic acid. The column was monitored at 280 nm (solid line) and by ninhydrin after alkaline hydrolysis (broken line). Fractions of 5 mL were collected at 20 mL/h.

as described by Koide et al. (1978). Six small peptides were thus isolated and found to be identical in composition with CB1, CB2, CB3, CB4, CB7, and CB8, previously described by Saari & Fischer (1973).

Except for CB15, CB17C, and CB18, which have been described in the two accompanying papers, the amino acid compositions of the other 12 cyanogen bromide fragments are listed in Table I.

Sequence Determination of the Cyanogen Bromide Fragments from Segment H_s. Of the 12 fragments listed in Table I, the sequences of the smallest, CB1, CB2, and CB3, have been determined (Cohen et al., 1973) and were not further examined. The sequences of the other nine were determined as described below.

CB4: This peptide (150 nmol) was coupled to TETA resin[1] through the carboxyl-terminal homoserine and analyzed in duplicate on the Sequemat (Table II) to yield the sequence Asp-Gly-Ala-Asn-Val-Glu-Hse, which is consistent with the partial sequence observed by Cohen et al. (1973).

CB7: Similar analysis of CB7 (150 nmol) yielded the sequence Leu-Asn-Gly-Ala-Leu-Thr-Ile-Gly-Thr-Hse. However, analysis of the first two residues was considered to be tentative, since the amino terminus was identified as the methylthiohydantoin, and the yield of the phenylthiohydantoin at the second turn was much lower than that of subsequent derivatives. This anomaly was not further examined because this portion of the sequence was consistent with that determined by Cohen et al. (1973) and a later experiment indicated that hydroxylamine cleaved an asparaginyl-glycine bond corresponding to this position in this fragment.

CB8: Sequencer analysis of CB8 (300 nmol) yielded the entire sequence, Ile-Gly-Gly-Lys-Ala-Ala-Pro-Gly-Tyr-His-Hse, except for the carboxyl-terminal homoserine which was placed by comparison of the composition and the sequence. The sixth residue, alanine, was also considered to be tentative, because of a lower yield than expected. Both assignments are in agreement with the partial sequence of Cohen et al. (1973).

CB9: This entire fragment (250 nmol) was also analyzed on the Sequencer, which identified 12 of the 13 residues, i.e., Ser-Leu-Val-Glu-Glu-Gly-Ala-Val-Lys-Arg-Ile-Asn-Hse. Once again the carboxyl-terminal homoserine was deduced from the composition.

CB10: During Sequencer analysis of CB10 (1 μmol), the fragment escaped from the reaction cup after the tenth turn, probably because the carboxyl-terminal portion is quite hydrophobic. Subsequently the peptide was coupled to TETA resin by its carboxyl-terminal homoserine and analyzed in duplicate (100 and 300 nmol) on the Sequemat. These data, together with the Sequencer analysis, yielded the sequence: Ala-Glu-Glu-Ala-Gly-Glu-Glu-Asn-Phe-Phe-Ile-Phe-Gly-Hse.

CB13: Since its first isolation by Saari & Fischer (1973), it has been assumed that this fragment is derived from the carboxyl-terminal portion of the whole molecule because it alone lacks homoserine. The intact fragment (1.4 μmol) was analyzed on the Sequencer (Table II) to yield the sequence of 40 of its 42 residues, although tryptophan 824 (Figure 3) in turn 25 and arginine-832 in turn 33 were only tentatively identified. These two residues were subsequently confirmed and the remainder of the sequence determined using tryptic peptides (Table III).

Approximately 3 μmol of CB13 was digested with Tos-PheCH_2Cl-trypsin for 1 h, acidified by addition of 0.1 volume of 88% formic acid, applied to a column of Sephadex G-25, and separated into six fractions (Figure 4A). Four peptides in fraction V were further separated by preparative high voltage paper electrophoresis at pH 3.7 and eluted with 10% acetic acid. Two peptides were similarly isolated from fraction I. Since they had identical compositions but different mobilities, it is probable that in one of them an amino-terminal glutamine cyclized to pyrrolidone carboxylic acid during the gel filtration. The two peptides were combined (T-8) and used for further study as described below.

T-9 was found to represent the carboxyl-terminal portion (residues 833 through 841) by analysis in the Sequemat after attachment of lysine-839 to DITC-3-aminopropyl glass (Table II). The amino terminus, Leu-Pro-Ala-Pro-Asp-Glu–, overlapped the analysis of cycles 34 through 39 of the intact fragment. Although no phenylthiohydantoins were observed after the sixth turn (glutamic acid-838), acid hydrolysis of the extract of the seventh turn yielded isoleucine and proline (1.00:0.85), suggesting the carboxyl-terminal sequence Glu-Lys(Ile,Pro).

A carboxyl-terminal proline in fragment CB13 was established by digestion of the intact fragment with carboxypeptidase Y, which released 0.82 equivalents of proline in 60 min (as well as 0.29 and 0.26 equiv of isoleucine and lysine, respectively). Similar digestion of peptide T-8 also indicated a carboxyl-terminal proline, but the ratio of the release of iso-

[1] Abbreviations used: TETA resin, triethylaminetetramine coupled to chloromethylpolystyrene resin; DITC, *p*-phenylene diisothiocarbamoyl; Tos-PheCH_2Cl-trypsin, trypsin (Worthington) treated with L-1-tosylamido-2-phenylethyl chloromethyl ketone; $NaDodSO_4$, sodium dodecyl sulfate.

TABLE I: Amino Acid Compositions[a] of CB Fragments[b] from Segment H_s.

	CB1	CB2	CB3	CB4	CB7	CB8	CB9	CB10	CB13	CB22	CB23	CB24
Residue No.	764-765	615-617	679-681	692-698	682-691	604-614	428-440	699-712	800-841	713-763	766-799	618-678
Figure	2-D	2-D	2-D	2-D	2-D	2-D	2-D	2-A	2-C	2-C	2-A	2-B
Fraction[c]	V-4	V-3	V-4	V-3	V-2	V-2	V-1	V[d]	III-2	III-1	IV	II-1
Amino Acid												
CmCys											0.50 (1)	
Asx				1.66 (2)	1.20 (1)		1.01 (1)	1.09 (1)	2.99 (3)	8.23 (8)	3.23 (3)	7.38 (7)
Thr					1.57 (2)				1.89 (2)		1.02 (1)	3.87 (4)
Ser							0.75 (1)		3.40 (4)	2.84 (3)	1.03 (1)	3.59 (4)
Hse	0.68 (1)	0.70 (1)	0.98 (1)	0.72 (1)	0.84 (1)	0.74 (1)	0.69 (1)	0.59 (1)		0.98 (1)	0.96 (1)	0.67 (1)
Glx				0.96 (1)			1.87 (2)	3.47 (4)	4.86 (5)	9.05 (9)	5.19 (5)	5.61 (5)
Pro						0.96 (1)			3.96 (4)	3.21 (3)	0.91 (1)	2.08 (2)
Gly				1.11 (1)	2.00 (2)	2.79 (3)	1.10 (1)	1.94 (2)	2.00 (2)	2.21 (2)		5.28 (5)
Ala		1.00 (1)		1.00 (1)	1.19 (1)	2.00 (2)	1.00 (1)	2.00 (2)	4.40 (4)	1.00 (1)	2.00 (2)	6.00 (6)
Val				0.85 (1)			1.53 (2)		1.84 (2)	2.85[e] (3)	2.81 (3)	6.75[e] (7)
Ile					0.87 (1)	1.34 (1)	0.90 (1)	0.97 (1)	4.10 (5)	3.99[e] (4)		6.34[e] (7)
Leu	1.00 (1)				1.57 (2)		0.98 (1)		1.13 (1)	4.00 (4)	1.25 (1)	4.98 (5)
Tyr						0.81 (1)			0.81 (1)	2.91 (3)	3.04 (3)	0.82 (1)
Phe			1.00 (1)					2.67 (3)	1.05 (1)	2.94 (3)	2.14 (2)	0.70 (1)
Lys		0.96 (1)	0.98 (1)			0.99 (1)	0.79 (1)		1.99 (2)	2.04 (2)	2.97 (3)	1.88 (2)
His						1.09 (1)					2.02 (2)	0.91 (1)
Arg							0.74 (1)		4.95 (5)	5.10(5)	3.89 (4)	2.57 (3)
Trp									[f] (1)		[f] (1)	
% Yield	50	35	40	38	35	33	60	85	65	60	85	30

[a] The numbers in parentheses indicate those found by sequence analysis. [b] CB17C, CB18, and CB15 are described in Koide et al. (1978) and Hermann et al. (1978). [c] CB1 and CB3, CB2 and CB4, and CB7 and CB8 were further separated from fractions V-4, V-3, and V-2 on an AG-1X2 column. [d] Precipitated (see text). [e] Value with 96-h acid hydrolysate. [f] Tryptophan was detected by Ehrlich's reagent (Smith, 1960).

leucine and lysine (0.51/0.59) was reversed from that observed above. However, it appears that T-8 may be a mixture of two peptides, one representing residues 831–841, and the other (resulting from cleavage of the Lys-Ile bond) consisting of residues 831–839. This would be consistent with both the low values of isoleucine and proline in the composition of T-8 (Table III) and the time course of amino acids released by carboxypeptidase Y. Thus the carboxyl-terminal sequence Lys-Ile-Pro is confirmed (Figure 3).

CB22: The specific strategy of the sequence determination of CB22 (51 residues) is illustrated in Figure 5. Sequencer analysis of the intact fragment (1.4 μmol) yielded the sequence of the amino-terminal 35 residues (Table II), although identifications of isoleucine 741 at turn 29 and serine-746 at turn 34 were only tentative. Since all five arginine residues were placed, the remaining sequence was established from lysine-containing tryptic peptides.

Approximately 4 μmol of the fragment was digested with Tos-PheCH_2Cl-trypsin for 3 h, acidified with 0.1 volume of 88% formic acid, applied to a column of Sephadex G-25, and separated into five fractions as shown in Figure 4B. Fractions I, II, III, and V each contained a single homogeneous peptide, i.e., T-5, T-1, T-4, and T-3, respectively (Table III). Three peptides, T-2, T-6, and T-7, were further separated from fraction IV by preparative high voltage paper electrophoresis at pH 3.7 (2000 V for 1.5 h). The compositions and electrophoretic mobilities (pH 6.5) of T-1, T-2, T-3, and T-4 agreed with those predicted for residues 713–719, 720–723, 724–733, and 724–738 in the amino-terminal sequence of CB22 (Figure 5). Sequencer analysis of T-4 (500 nmol) confirmed the sequences of residues 724–737 (Table II).

The sequence of the carboxyl-terminal half of CB22 was determined with the aid of three tryptic peptides, T-5, T-6, and T-7, and their thermolytic subdigestion products. Each tryptic peptide was analyzed in the Sequemat after attachment of the carboxyl-terminal lysine or homoserine to DITC-3-aminopropyl glass or TETA resin (Table II). The amino-terminal methylthiohydantoins of T-5 and T-6 were not identified. In the analysis of T-5, serines at turns 7, 8, and 12 were not clearly identified, but the partial sequence clearly overlapped residues 740–747. To complete the sequence of T-5, 1 μmol of the peptide was digested with thermolysin at 37 °C in 0.1 M NH_4HCO_3, pH 8.0 for 3 h, and three products were separated on an AG 50W-X2 column (Table III). The composition of T-5-Th-1 was consistent with the sequence of residues 739–743, whereas that of T-5-Th-2 confirmed the placement of seryl residues at residues 745 and 746. Digestion of T-5-Th-3 with carboxypeptidase Y provided the carboxyl-terminal sequence of T-5 (0.10 proline and 0.47 lysine were released in 30 min, and 0.26 proline and 0.95 lysine in 2 h at 37 °C in 0.1 M pyridine–acetate, pH 6.0).

Sequemat analysis of T-6 yielded phenylthiohydantoins in much lower yield than expected (only 5%). Apparently the original peptide contained an amino-terminal glutamine, most of which cyclized to a pyrollidone carboxylic residue, but a small amount (5%) lost its amide to form amino-terminal

TABLE II: Automated Edman Degradations of Fragments Derived from Residues 265 through 841.

Fragment analyzed (residues no.)	Amount applied (nmol)	Techniques used: Degradation[a] Identification[b]	Continuous sequence proven	Identification tentative	Identification poor or lacking
H_s (265-841)	95	A GL	265-274	267,268	269,273
AC-5 (265-319)	50	A GL	265-267		
HA-2 (265-452)	365	A GL	265-283	281,283	276,277
AC-3 (265-632)	50	A GL	265-267		
AC-4 (320-632)	450	A GL,HP	320-354		322,349,351,352,353
AC4-TS2 (427-437)	90	B* HP	427-435		
CB9 (428-440)	250	B GL	428-439		
AC4-TS3 (438-456)	120	B* GL,HP	438-447	444	
HA-1 (484-684)	75	A GL	484-486		
AC4-TSS4 (601-632)	100	B GL,HP	601-627		601,623
CB8 (604-614)	300	B GL	604-613	609	
CB24 (618-678)	1250	B GL	618-669	630,660,666	662
"	100	B HP	618-645		631,638,640
AC-2 (633-841)	50	A GL	633-635		
AC2-TS1 (649-713)	250	B* GL,HP	649-682	667	
CB24-TC (655-678)	200	C GL,HP	655-678		655
CB7 (682-691)	150	C GL	682-691	683	
HA-3 (684-841)	300	A GL	684-703	700	690,692,697,701
CB4 (692-698)	150	C GL	692-698		
"	150	C GL	692-698		
CB10 (699-712)	1000	B GL	699-709	709	
"	100	C GL	699-712	709	
"	300	C GL	699-712		707,708
AC2-TS1-Th1 (709-713)	40	B HP	709-712		
CB22 (713-763)	1400	B GL	713-747	741,746	
CB22-T4 (724-738)	500	B GL	724-738	738	
CB22-T5 (739-752)	200	C GL,HP	739-750	746	739,745,750
CB22-T6 (753-758)	200	C GL	753-757	757	753
CB22-T7 (759-763)	200	C GL	759-763		
AC2-TS2-Th1 (764-769)	25	B HP	764-768		766,767
CB23 (766-799)	800	B GL	766-796	778,782,796	787,794
CB23-T5 (786-791)	200	C GL	786-790		786
CB23-T7 (795-798)	200	B GL	795-798		
CB13 (800-841)	1400	B GL	800-839	832	824
CB13-T9 (833-841)	100	C GL	833-838		833

[a] Edman degradation by (A) liquid phase (Sequencer), protein program, (B) liquid phase (Sequencer), peptide program, where * indicates that polybrene was included as an inert carrier, or (C) solid phase (Sequemat). [b] Identification by gas–liquid chromatography (GL) and spot tests for histidine and arginine or high performance liquid chromatography (HP).

glutamic acid. Thus, both components would be acidic, as observed at pH 6.5; they would copurify, but only one would be susceptible to sequence analysis. This was confirmed by thermolytic subdigestion of T-6 (1 μmol) which yielded a tetrapeptide (T-6-Th-1) which did *not* bind to AG 50W-X2 at pH 2 and did *not* react with ninhydrin. Digestion of this peptide with carboxypeptidase Y confirmed the sequence Gln-Pro-Asp-Leu in T-6 (0.06 aspartic acid and 0.92 leucine released in 30 min and 0.19 aspartic acid and 1.00 leucine in 2 h). A second thermolytic peptide T-6-Th-2 provided the carboxyl-terminal Phe-Lys.

The sequence of peptide T-7 was determined with the Sequemat after attachment through its homoseryl residue.

Although there is no direct evidence for overlaps between T-5, T-6, and T-7, it is evident that T-7 contains the carboxyl-terminal homoserine of CB22, that T-5 extends the amino-terminal sequence of CB22 to lysine-40 and that T-6 can only occupy residues 41-46, as shown in Figure 5.

CB23: Sequencer analysis of the intact fragment (800 nmol) yielded the sequence of 31 of its 34 residues (Table II), although identifications of glutamic acid-778, cysteine-782, serine-787, arginine-794, and tryptophan-796 (Figure 3) at turns 13, 17, 22, 29, and 31 were only tentative. These residues were more clearly identified and others confirmed using tryptic peptides.

Approximately 1.6 μmol of CB23 was digested with Tos-PheCH_2Cl-trypsin for 20 h. The lyophilized digest was dissolved in 5 mL of 9% formic acid and separated on a column of Sephadex G-25 into seven fractions (Figure 4C). Each fraction contained a single homogeneous peptide (Table III)

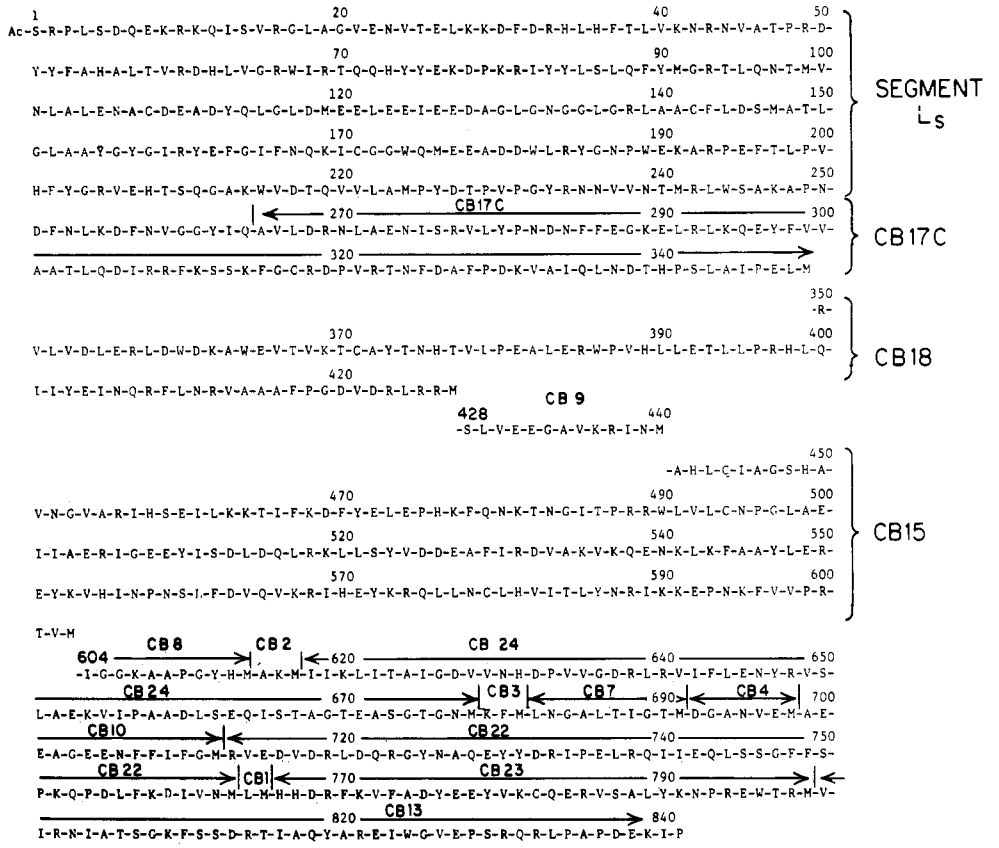

FIGURE 3: The amino acid sequence of rabbit muscle glycogen phosphorylase *b*, broken at the four points where overlapping structures are provided to align previously reported data. Proof of the sequence of residues 1-349 was reported by Koide et al. (1978) and of residues 350-427 (CB18) and 441-603 (CB15) by Hermann et al. (1978). The sequences of the short peptide CB9 (residues 428-440), the carboxyl-terminal region (residues 604-841), and overlapping structures uniting the broken segments are documented in this communication. Cyanogen bromide fragments are identified only from residue 265-841. The sequence is given in one-letter code: A (alanine), C (cysteine), D (aspartic acid), E (glutamic acid), F (phenylalanine), G (glycine), H (histidine), I (isoleucine), K (lysine), L (leucine), M (methionine), N (asparagine), P (proline), Q (glutamine), R (arginine), S (serine), T (threonine), V (valine), W (tryptophan), and Y (tyrosine).

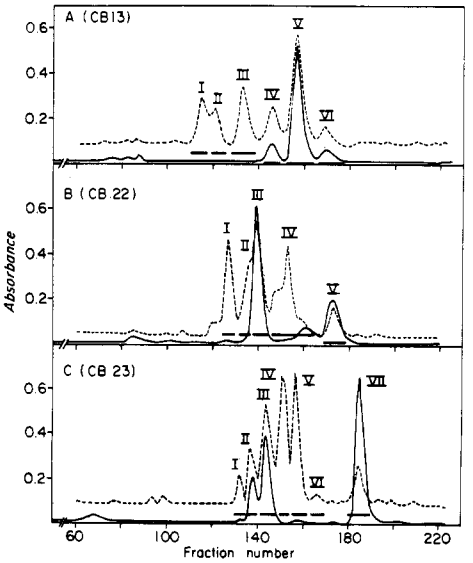

FIGURE 4: Separation of tryptic peptides of (A) 3.0 μmol of CB13, (B) 4.0 μmol of CB22, and (C) 1.6 μmol of CB23 on a column (2.5 × 200 cm) of Sephadex G-25 Superfine with 9% formic acid. Fractions of 5 mL were collected at 25 mL/h and pooled as indicated. The effluents were monitored at 280 nm (solid line) and by ninhydrin after alkaline hydrolysis (broken line).

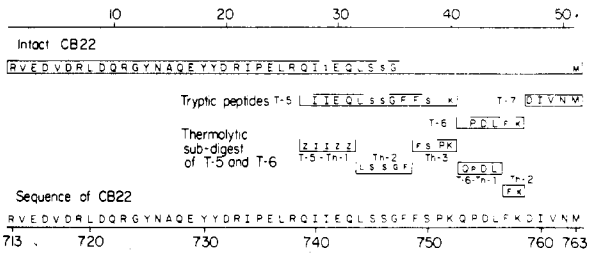

FIGURE 5: Strategy of sequence analysis of CB22. The 51 residues (upper numbers) are placed in the sequence of phosphorylase (lower numbers) in one-letter code (Figure 3). Z indicates glutamic acid or glutamine. Each peptide is symbolized by a horizontal bar; upper enclosure of that bar indicates the portion of sequence proven in that peptide; small capital letters indicate that a residue is either tentatively identified (Table II) or placed by reason of enzyme specificity or composition (see text).

except fraction IV which was further separated into 2 peptides by preparative high voltage paper electrophoresis at pH 3.7 (2000 V for 1.5 h). Although the analytical value of carboxymethylcysteine indicated partial oxidative loss, its presence in peptide T-4 (residues 782-785, Figure 3) as residue 782 is in accord with the tentative identification of the phenylthiohydantoin of dehydroserine (after silylation) at this position (residue 17) in the Sequencer analysis of the intact fragment. Serine-787 was unambiguously identified by the composition of T-5 (residues 786-791) and by analysis of 200 nmol of T-5 in the Sequemat after attachment of the carboxyl-terminal lysine (Table II). Arginine-794 was confirmed by the composition of T-6. Sequencer analysis of T-7 (200 nmol) confirmed glutamic acid-795 and tryptophan-796 (the last residues identified in the analysis of the intact fragment) and ex-

TABLE III: Amino Acid Compositions[a] of Tryptic Peptides from Fragments CB13, CB22, and CB23.

Parent Fragment	CB13						CB22										CB23							
Peptide	T-1	T-2	T-3	T-4	T-5	T-6	T-7	T-8	T-9	T-1	T-2	T-3	T-4	T-5	T-5-Th-1	T-5-Th-2	T-5-Th-3	T-6	T-6-Th-1	T-6-Th-2	T-7	T-1	T-2	T-3	T-4	T-5	T-6	T-7	T-8
Residues No.[b]	660–662	803–809	810–814	815–819	815–821	820–821	822–830	831–841	831–841	713–719	720–723	724–733	724–738	739–752	739–744	744–748	748–752	753–758	753–756	757–758	759–763	764–769	770–771	772–781	782–785	786–791	792–794	795–798	799
Fraction No.	A-V[c]	A-III	A-V[c]	A-VI	A-IV	A-V[c]	A-V[c]	A-I	A-II	B-II	B-IV[d]	B-V	B-III	B-I	(see the text)						B-IV[d]	(see the text)	C-I	C-V	C-III	C-IV[d]	C-II	C-IV[d]	C-VII	C-VI
Fig. 4																													
Amino Acid																													
Cys																													
Asx		1.00 (1)	1.14 (1)					1.05 (1)	1.10 (1)	2.05 (2)	1.06 (1)	1.87 (2)	2.43 (2)	0.37 (0)				1.10 (1)	1.04 (1)		2.00 (2)	1.10 (1)		1.05 (1)			1.07 (1)		
Thr		0.90 (1)		0.90 (1)	0.98 (1)																							0.94 (1)	
Ser		0.89 (1)	1.90 (2)				0.94 (1)																			0.85 (1)			
Asn																					0.60 (1)								1.00 (1)
Glx				1.21 (1)	1.30 (1)		1.74 (2)	2.02 (2)	1.28 (1)	1.41 (1)	1.03 (1)	2.06 (2)	2.45 (3)	2.27 (3)	3.00 (3)			1.01 (1)	0.95 (1)				2.17 (2)	2.06 (2)				0.99 (1)	
Pro							1.06 (1)	2.59 (3)	2.20 (3)				1.03 (1)	1.07 (1)			1.00 (1)	0.94 (1)	0.94 (1)						0.96 (2)			
Gly		0.98 (1)					1.05 (1)						1.00 (1)	0.96 (1)															
Ala		1.00 (1)		1.00 (1)	2.00 (2)	1.00 (1)		1.00 (1)	1.00 (1)			1.11 (1)	1.00 (1)				1.23 (1)						1.00 (1)			0.94 (1)			
Val	0.58 (1)						1.00 (1)			1.70 (2)																			
Ile	0.63 (1)	0.93 (1)		1.42 (1)	1.24 (1)		1.05 (1)	0.70 (1)	0.70 (1)				0.99 (1)	1.21 (2)	1.70[e] (2)						0.85 (1)			1.99 (2)		1.04 (1)			
Leu								1.00 (1)	0.96 (1)		1.00 (1)		1.05 (1)	1.19 (1)	0.87 (1)				1.01 (1)	1.00 (1)	0.71 (1)					1.00 (1)			
Tyr				0.80 (1)	0.67 (1)							2.61 (3)	2.27 (3)											1.98 (2)		1.07 (1)			
Phe			1.00 (1)											1.74 (2)	1.00 (1)	0.84 (1)	1.00 (1)			0.88 (1)			1.00 (1)	1.00 (1)					
Lys		0.97 (1)						1.00 (1)	0.97 (1)					1.00 (1)		1.00 (1)	1.00 (1)			1.00 (1)			1.06 (2)	1.00 (1)					
His																						1.90 (2)							
Arg	1.00 (1)		1.22 (1)		1.37 (1)	1.16 (1)	1.17 (1)	1.04 (1)		2.00 (2)	0.79 (1)	1.08 (1)	2.27 (2)	0.26 (0)								1.00 (1)			1.00 (1)		1.00 (1)	1.00 (1)	
Trp							1 (1)																					B.L. (1)	
Σ Yield	37	77	32	30	46	27	45	37	27	63	15	34	46	97	20	20	27	11	5	5	10	21	57	55	38	55	34	58	9

[a] Compositions of 24-h acid hydrolysates. Numbers in parentheses are those found by sequence analysis (Figure 3). [b] In the complete sequence (Figure 3). [c] Four peptides in fraction V (CB13) were separated by paper electrophoresis at pH 3.7. [d] Three peptides in fraction IV (CB22) were separated by paper electrophoresis at pH 3.7. [e] Two peptides in fraction IV (CB23) were separated by electrophoresis at pH 3.7. [f] Values in 96-h acid hydrolysates. [g] Tryptophan was detected by Ehrlich's reagent (Smith, 1960).

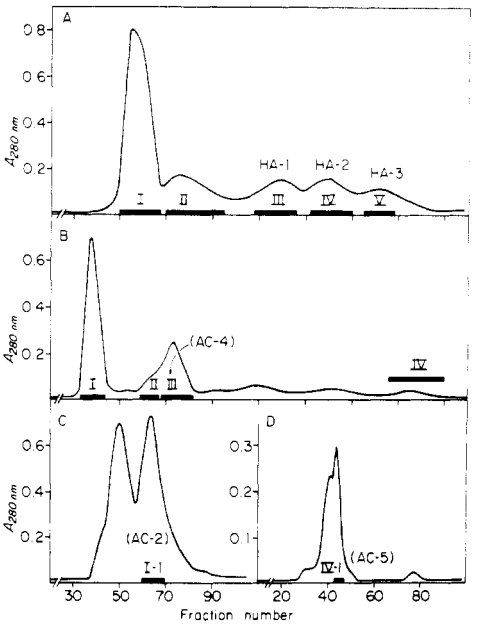

FIGURE 6: (A) Fragments generated from segment H_s (170 mg) by cleavage with hydroxylamine were separated on a column (5.0 × 145 cm) of Sephadex G-100 Fine in 9% formic acid. Fractions of 9 mL were collected at 36 mL/h and pooled as indicated by the horizontal bars. (B) Separation of fragments generated from segment H_s (140 mg) by mild acid cleavage (pH 2.5). Conditions of chromatography were identical with A, except that fractions of 20 mL were collected. (C) Purification of fraction I (35 mg) from B on a column (1.5 × 85 cm) of Sephadex G-200 Fine in 0.5 M NH_4HCO_3, 6 M guanidine HCl, pH 7.8. Fractions of 1.5 mL were collected at 3 mL/h, pooled as indicated by the horizontal bar, and desalted by dialysis against water. (D) Purification of fraction IV (15 mg) from B on a column (2.5 × 110 cm) of Sephadex G-50 Superfine in 9% formic acid. Fractions of 5.0 mL were collected at 20 mL/h.

tended the sequence to arginine-798 which must be the residue penultimate to the carboxyl-terminal homoserine, since all of the amino acid residues in the fragment were recovered in eight tryptic peptides and one of these (T-8) was free homoserine.

CB24: This fragment contains 61 residues including 3 arginyl and 2 lysyl residues, all five of which were identified in 52 cycles of analysis in the Sequencer (Table II). The carboxyl-terminal portion of CB24 was selectively isolated from a tryptic digest of 200 nmol of the fragment by the method of Horn & Laursen (1973). Sequemat analysis (CB24-TC in Table II) yielded a sequence which overlapped the amino-terminal segment by 15 residues; it also confirmed aspartic acid-660 and serine-666 (Figure 3) and provided the sequence of the carboxyl-terminal nonapeptide including the homoseryl residue.

Cleavage of Segment H_s at Asparaginylglycine Bonds. Since three Asn-Gly bonds within segment H_s had been identified by structural analysis of fragments CB15 and CB7, 210 mg (approximately 3.2 μmol) of carboxymethylated H_s was cleaved with hydroxylamine as described by Koide et al. (1978). The reaction mixture was desalted by dialysis against water and lyophilized to yield 140–180 mg in successive experiments. $NaDodSO_4$ gel electrophoresis indicated a complex mixture of about ten large fragments. The products (170 mg) were dissolved in 25 mL of 9% formic acid and applied to a column (5 × 145 cm) of Sephadex G-100 Fine in 9% formic acid. A typical elution profile is shown in Figure 6A. Fractions I and II, accounting for approximately 50% of the products by weight, contained three major large fragments (including intact H_s), which, by gel electrophoresis and Sequencer analysis, appeared to be overlapping fragments generated by incomplete cleavage. They were not further purified. Similar examination

TABLE IV: Amino Acid Compositions[a] of HA Fragments.

residue no.:	HA-2 265–452	HA-1 484–683	HA-3 684–841	Σ[b]	H_s[c] 265–841	CB23-13 isolated from HA-3 766–841
amino acid						
CM-Cys	1.2 (3)	1.4 (2)	0.4 (1)	3.0	(6.6)	0.5 (1)
Asx	23.1 (23)	20.0 (21)	16.1 (17)	59.2	(63.1)	6.8 (6)
Thr	8.5 (8)	6.6 (7)	5.3 (5)	20.4	(23.0)	2.9 (3)
Ser	7.9 (6)	6.1 (7)	7.1 (8)	21.1	(21.7)	4.0 (5)
Glx	20.7 (18)	19.3 (19)	22.5 (24)	62.5	(65.1)	11.4 (10)
Pro	10.4 (9)	7.8 (8)	7.6 (8)	25.8	(27.5)	4.9 (5)
Gly	7.9 (5)	8.4 (11)	9.0 (9)	25.3	(28.3)	3.1 (2)
Ala	17.0 (17)	15.0 (15)	11.0 (11)	43.0	(44.0)	6.0 (6)
Val	15.7 (17)	14.9 (18)	8.6 (9)	39.2	(42.1)	6.2 (5)
Met	3.5 (3)	4.4 (5)	4.7 (6)	12.6	(13.5)	0.5 (1)
Ile	10.4 (9)	12.5 (18)	8.5 (11)	31.4	(41.9)	4.4 (5)
Leu	22.0 (23)	17.5 (20)	9.9 (8)	49.4	(56.7)	2.6 (2)
Tyr	2.9 (4)	5.1 (8)	4.4 (7)	12.4	(19.5)	2.2 (4)
Phe	8.9 (9)	7.2 (6)	8.1 (9)	24.2	(30.3)	4.0 (3)
Lys	9.7 (8)	14.8 (16)	6.7 (7)	31.2	(35.8)	2.1 (2)
His	6.4 (6)	5.4 (5)	2.9 (2)	14.7	(15.0)	4.8 (5)
Arg	15.8 (17)	12.4 (13)	12.7 (14)	40.4	(45.2)	9.0 (9)
Trp	ND (3)	ND (1)	ND (2)	ND	(5.2)	ND (2)
no. of residues	(188)	(200)	(158)			(76)
% yield	29	24	26			13
approx. mol wt[d]	22 000	24 000	18 000	64 000	68 000	ND

[a] Compositions of 24-h acid hydrolysates are calculated on the basis of 17, 15, 11, and 6 residues of alanine in HA-2, HA-1, HA-3, and CB23-13. Numbers in parentheses are from the sequence of constituent cyanogen bromide fragments (Table V). [b] The sum of values for HA-1, -2, and -3. [c] Taken from Koide et al. (1978). [d] By $NaDodSO_4$ disc gel electrophoresis.

TABLE V: Fragments Isolated[a] after Cyanogen Bromide Cleavage of HA- and AC- Fragments.

PARENT HA-fragment	yield (μmol)	Generated only from HA-	Generated from both HA- and AC-	Generated only from AC-	yield (μmol)	PARENT AC-fragment
HA-2 (0.93 μmole)	0.15	CB17C		CB17Cac	0.14	AC-4 (0.46 μmole)
	0.15		CB18		0.14	
	0.25		CB9		0.14	
	0.25	CB15N		CB15	0.27	
HA-1 (0.77 μmole)	0.15	CB15C				
	0.24		CB8		0.20	
	0.22		CB2		0.23	
	0.18	CB24		CB24N	0.28	
				CB24C	0.04	AC-2 (0.48 μmole)
	0.15		CB3		0.08	
	0.20	CB7N		CB7	0.20	
HA-3 (0.83 μmole)	0.25	CB7C				
	0.25		CB4		0.13	
	0.25		CB10		0.12	
	0.09		CB22		0.04	
	0.25		CB1		0.07	
	0.15		CB23		0.22	
	0.09		CB13		0.04	
	0.11	CB23-13				

[a] This table is structured to draw attention to the CB fragments common to two parent fragments (central column) and to those derived exclusively from a fragment HA- (left) or a fragment AC- (right).

of fractions IV and V revealed essentially one fragment in each, which were designated HA-2 and HA-3, respectively. Fraction III contained, even after rechromatography, not only a major fragment (HA-1), but also two minor fragments (5–10% of the major fragment) with slightly higher molecular weights. Molecular weight and amino-terminal sequence determinations indicated that these minor fragments resulted from incomplete cleavage of the two asparaginylglycine bonds within CB15 (Hermann et al., 1978). The amino acid compositions of fractions HA-1, HA-2, and HA-3 are shown in Table IV, and the extent of their amino-terminal sequence analyses is shown in Table II.

Fraction HA-2 was shown by Sequencer analysis (Table II) to be derived from the amino terminus of segment H_s (Figure 1), beginning with alanine-265, as described by Koide et al. (1978). Similarly, the amino-terminal sequence of fraction HA-1 corresponded to that of the interior of fragment CB15 (Hermann et al., 1978) beginning with glycine 484. Sequencer analysis of 20 residues of HA-3 (Table II) indicated that this fragment began with glycine 684, the third residue in CB7, and continued through CB4 into CB10 (5 residues).

The compositions of HA-1, HA-2, and HA-3 were further clarified by cyanogen bromide treatment and isolation of their constituent fragments as described above, but using a smaller column (2.5 × 90 cm) of Sephadex G-50. The fragments isolated are listed in Table V and Figure 1. Thus, HA-2 contains, besides the carboxyl-terminal portion of CB17, fragments CB18 and CB9 (in unknown order) and a fragment (CB15N) corresponding in composition to that segment leading to the first asparaginylglycine bond in CB15 (Hermann et al., 1978). The portion of the sequence between the two asparaginyl bonds of CB15 (residues 453–483, Hermann et al., 1978) was apparently lost during dialysis of the larger fragments. Fragment HA-1 continues from residue 484 through the remainder of CB15 and includes four other small fragments (in unknown order) and the amino-terminal Leu-Asn of CB7. Cleavage of Asn-Gly in CB7 separated HA-1 from HA-3. Beyond the alignment CB7-CB4-CB10 in HA-3 are found CB22 and CB1 (in unknown order). CB23 is placed next to the carboxyl-terminal fragment CB13 on the basis of the composition of a

TABLE VI: Amino Acid Compositions[a] of AC Fragments.

fragment:	AC-5	AC-3	AC-4	AC-2	Ac-5 + AC-4 + AC-2	AC-3 + AC-2	H_s[b]
residue no.:	265-319	265-632	320-632	633-841			265-841
amino acid							
CM-Cys	0.6 (1)	2.7 (5)	2.4 (4)	1.1 (1)	4.1	3.8	6.6
Asx	8.2 (8)	42.1 (42)	36.7 (34)	22.0 (22)	66.9	64.1	63.1
Thr	1.6 (1)	14.4 (14)	15.0 (13)	7.9 (8)	24.5	22.3	23.0
Ser	2.9 (3)	9.8 (10)	7.6 (7)	11.3 (12)	21.8	21.1	21.7
Glx	6.8 (6)	36.3 (36)	29.7 (30)	28.7 (29)	65.2	65.0	65.1
Pro	1.7 (1)	17.4 (16)	14.9 (15)	10.4 (10)	27.0	27.8	27.5
Gly	2.1 (2)	14.3 (13)	11.9 (11)	13.1 (13)	27.1	27.4	28.3
Ala	4.0 (4)	28.0 (28)	24.0 (24)	16.0 (16)	44.0	44.0	44.0
Val	3.6 (4)	29.1 (31)	24.9 (27)	12.6 (14)	41.1	41.7	42.1
Met	0 (0)	5.6 (6)	6.8 (6)	5.6 (8)	12.4	11.2	13.5
Ile	2.7 (2)	24.2 (27)	22.2 (25)	12.1 (14)	37.0	36.3	41.9
Leu	6.2 (6)	39.5 (40)	34.8 (34)	13.8 (13)	54.8	53.3	56.7
Tyr	1.7 (2)	11.0 (12)	8.4 (10)	6.6 (8)	16.7	17.6	19.5
Phe	4.9 (5)	14.9 (16)	11.6 (11)	10.4 (11)	26.9	25.3	30.3
Lys	4.0 (4)	26.4 (27)	24.0 (23)	8.5 (9)	36.5	34.9	35.8
His	0 (0)	11.3 (13)	13.1 (13)	2.6 (2)	15.7	13.9	15.0
Arg	5.8 (6)	26.8 (28)	21.7 (22)	16.2 (17)	43.7	43.0	45.2
Trp	N.D. (0)	N.D. (4)	N.D. (4)	N.D. (2)	N.D.	N.D.	5.2
no. of residues	56.8 (55)	353.8 (368)	309.7 (313)	198.9 (209)	565.4	552.7	584.5
% yield	60	23	46	51			
approx. mol. wt.[c]	6 000	42 000	36 000	24 000	66 000	66 000	68 000
amino-terminal sequence	Ala-Val-Leu- - -	Ala-Val-Leu- - -	Pro-Val-Arg- - -	Pro-Val-Val			Ala-Val-Leu- - -

[a] Compositions of 24-h acid hydrolysates are calculated on the basis of 4, 28, 24 and 16 residues of alanine per mole of fragment AC-5, AC-3, AC-4 and AC-2. Numbers in parentheses are from the constituent cyanogen bromide fragments (Table V). [b] Taken from Koide et al. (1978). [c] By $NaDodSO_4$ disc gel electrophoresis.

fragment (CB23-13), obtained in 13% yield, which appears to overlap these two fragments (Table IV).

Cleavage of Segment H_s at Aspartylproline Bonds. Since structural analysis of CB17 and CB24 had identified two Asp-Pro bonds in segment H_s, this segment was cleaved by mild acid treatment to yield three primary fragments AC-5, AC-4, and AC-2 (Figure 1).

The S-carboxymethylated heavy segment H_s (140 mg, ca. 2.1 μmol) was incubated in 70 mL of 10% acetic acid, containing 7 M guanidine hydrochloride (adjusted to pH 2.5 with pyridine), at 37 °C for 5 days. The reaction mixture was then desalted on a column (5 × 45 cm) of Sephadex G-25 Fine in 9% formic acid and lyophilized to yield approximately 140 mg of products. $NaDodSO_4$ disc gel electrophoresis demonstrated four major fragments and many trace components (ca. 1–5%), indicating that the cleavage had not been entirely restricted to Asp-Pro bonds, which would have generated at most 3 primary and 3 overlapping fragments.

The products were dissolved in 2 mL of 88% formic acid, diluted with 9 volumes of water and a small amount of precipitate removed by centrifugation. The acid-soluble portion was fractionated on a column (5 × 145 cm) of Sephadex G-100 Fine in 9% formic acid as shown in Figure 6B. Most of the material was recovered in four fractions, i.e., fractions I (50 mg), II (20 mg), III (35 mg), and IV (15 mg). Fraction I, containing mostly AC-2 (ca. 24 000 daltons) and several minor fragments of molecular weights greater than 50 000, was further fractionated on a column (1.5 × 85 cm) of Sephadex G-200 Fine in 0.5 M NH_4HCO_3, 6 M guanidine hydrochloride, pH 7.8 (Figure 6C) to yield pure fragment AC-2. Two major fragments (AC-3 and AC-4) in fractions II and III were partially separated by rechromatography on the G-100 column, but each was still contaminated by ca. 20% of the other. Another major fragment (AC-5) was isolated from fraction IV on a column (2.5 × 110 cm) of Sephadex G-50 Superfine in 9% formic acid (Figure 6D). Characteristics of the four major fragments thus obtained are listed in Table VI.

It is apparent from the amino-terminal sequences (Table II) and the sizes of the fragments (Table VI) that AC-3 overlaps AC-5 and AC-4, that AC-5 is derived from the amino-terminal region of segment H_s (residues 265–319), and that AC-4 is derived by cleavage of the Asp-Pro bond, 319–320 in CB17C (Koide et al., 1978). Since the overlapping fragment AC-3 will not be considered further, it is not included in Figure 1.

Fragment AC-2 appears to be derived from the carboxyl-terminal portion of the heavy segment H_s, since the sum of the compositions and molecular weights of AC-5, AC-4, and AC-2 agrees reasonably well with that of segment H_s and AC-5 and AC-4 have already been placed.

The two major fragments AC-4 and AC-2 were further characterized by treatment of each (0.5 μmol) with cyanogen bromide and isolation of the resulting fragments as described above [but on a smaller column (1.5 × 85 cm) of Sephadex G-50]. Each isolated fragment (see Table V) was identified by amino acid analysis. Thus, AC-4 includes the carboxyl-terminal region of CB17 (designated CB17Cac in Figure 1 and Table V), proceeds through 5 more methionyl residues, and ends with 15 residues at the amino terminus of CB24 (designated CB24N). Extended Sequencer analysis of AC-4 (Table II) confirmed proline-320 through leucine-348 and continued with the partial sequence X-Arg-X-X-X-Asp (Table II) which

TABLE VII: Amino Acid Compositions[a] of Methionine-Containing Peptides Isolated from HA-3, AC-4, or AC-2.

Peptide	HA3-T1	HA3-T2	AC4-TSS1	AC4-TS2	AC4-TS3	AC4-TSS4	AC2-TS1	AC2-TS1-Th1[b]	AC2-TS2	AC2-TS2-Th1[b]
Residue No.	684-713	799-802	348-350	427-437	438-456	601-632	649-713	709-713	739-769	764-769
Figure			7-B	7-A	7-A	7-B	7-C		7-C	
Fraction			I-2	II	II	I-1	I		II	
Amino Acid										
CmCys					0.28 (1)					
Asx	3.21 (3)				2.19 (2)	3.72 (3)	6.56 (6)		3.67 (4)	1.13 (1)
Thr	1.75 (2)					2.29 (2)	5.07 (5)		0.57 (0)	
Ser	0.59 (0)			0.95 (1)	1.11 (1)	0.48 (0)	4.30 (4)		2.12 (3)	
Glx	5.16 (5)			2.40 (2)		0.74 (0)	9.90 (9)		3.82 (4)	
Pro	0.28 (0)					1.71 (1)	1.23 (1)		1.74 (2)	
Gly	4.34 (5)			1.02 (1)	2.13 (2)	3.96 (4)	8.28 (8)	1.23 (1)	1.50 (1)	
Ala	4.00 (4)			1.00 (1)	4.00 (4)	4.00 (4)	9.00 (9)		0.79 (0)	
Val	1.34 (1)	1.21 (1)		1.82 (2)	1.72 (2)	2.52 (3)	2.88 (3)		1.01 (1)	
Met	2.34 (3)	0.63 (1)	0.71 (1)	0.73 (1)	0.70 (1)	2.28 (3)	4.88 (5)	0.63 (1)[c]	1.82 (2)	1.02 (1)[c]
Ile	2.01 (2)	1.00 (1)			1.92 (2)	4.38 (5)	4.15 (4)	0.91 (1)	2.07 (3)	
Leu	1.51 (1)		1.00 (1)	1.14 (1)	1.13 (1)	1.57 (1)	5.08 (4)		3.00 (3)	1.00 (1)
Tyr	0.34 (0)					0.89 (1)				
Phe	2.76 (3)					0.40 (0)	4.48 (4)	1.00 (1)	2.75 (3)	
Lys	0.32 (0)			0.96 (1)		2.75 (3)	1.79 (2)		1.56 (2)	
His					1.35 (2)	1.78 (2)			1.38 (2)	1.75 (2)
Arg	1.03 (1)	1.00 (1)	0.98 (1)	0.96 (1)	0.67 (1)		1.00 (1)	0.89 (1)	0.94 (1)	1.18 (1)
No. of Residues	30	4	3	11	19	32	65	5	31	6
% Yield	28	5	5	12	10	16	30	15	25	15

[a] Compositions of 24-h acid hydrolysates are calculated as ratios to the underlined number in each column. Numbers in parentheses are those found by sequence analysis (Figure 3). [b] 300 nmol of AC2-TS1 or AC2-TS2 was oxidized with performic acid and then digested with thermolysin (0.4 mg of enzyme in 0.2 mL of 0.1 M NH_4HCO_3, pH 8, for 2 h). Each digest was separated on preparative high-voltage paper electrophoresis at pH 6.5. In each case a single peptide containing arginine was detected with phenanthrenequinone and eluted with 10% acetic acid. [c] Estimated as methionine sulfone.

tentatively places CB18 after CB17. Similarly, AC-2 begins with proline-633 in CB24 and proceeds through regions giving rise to 8 other cyanogen bromide fragments including the carboxyl-terminal CB13.

The distribution of CB fragments between AC-4 and AC-2 is different from that between HA-1 and HA-3 (Figure 1). Analysis of AC-4 showed that CB8 and CB2 must precede CB24. Thus, the fragments in HA-1 can be placed in the order CB15C-(CB8, CB2)-CB24-CB3-CB7N. Since CB3 is found in AC-2 but not in HA-3, the order in AC-2 must be CB24C-CB3-CB7-CB4-CB10-(CB22, CB1)-CB23-CB13.

Alignment of Fragments Using Methionine-Containing Peptides from Enzymatic Digests. At this point the 15 cyanogen bromide fragments of segment H_s are partially aligned in the following order (where dashes indicate that an overlapping sequence has been observed, commas indicate that an overlap is lacking and parentheses enclose groups of contiguous fragments): CB17C-CB18, CB9, CB15, (CB8, CB2), CB24, CB3, CB7-CB4-CB10, (CB22, CB1), CB23-CB13. To complete this alignment eight methionine-containing peptides were isolated from enzymatic subdigests of fragments AC-4, AC-2, and HA-3.

Fragment AC-4 (90 mg) was succinylated as described by Koide et al. (1978) and the product desalted by dialysis. The succinylated fragment (80 mg) was then treated with 2 mg of Tos-PheCH_2Cl-trypsin in a pH-stat (pH 8) for 2 h at 37 °C and applied directly to a column of Sephadex G-50 (Figure 7A). Amino acid analysis of fractions I and II indicated the presence of methionine-containing peptides, two of which were isolated from fraction II by high voltage paper electrophoresis at pH 6.5 (AC4-TS2 and AC4-TS3 in Table VII). Since fraction I (26 mg), a mixture of several large peptides, could not be further resolved, it was subdigested with staphylococcal protease in 0.1 M NH_4HCO_3, pH 8.0, for 24 h and reapplied to the same column (Figure 7B). Methionine was detected by amino acid analysis in fractions I-1 and I-2. Fraction I-1 contained a peptide, AC4-TSS4, which resembled in composition (Table VII) residues 601 through 632. Fraction I-2 was further purified by preparative high voltage paper electrophoresis at pH 6.5 (2000 V; 90 min) to yield a homogeneous tripeptide AC4-TSS1 (Table VII).

Fragment AC-2 (60 mg) was similarly succinylated and digested with trypsin. The components were separated on a Sephadex G-50 column (Figure 7C) and methionine detected by amino acid analysis mainly in fractions I, II, and III. Fraction I comprised a homogeneous peptide AC2-TS1 containing 5 methionyl residues (Table VII). Fraction II consisted of another peptide (AC2-TS2) containing 2 methionyl residues. Fraction III was not satisfactorily resolved.

Fragment HA-3 (40 mg) was digested with Tos-PheCH_2Cl-trypsin for 20 h. The lyophilized digest was resuspended in 5 mL of 9% formic acid and incubated at 5 °C for 20 h. A precipitate formed having a composition (HA3-T1 in Table VII) consistent with the amino-terminal 30 residues of HA-3, including three methionyl residues. The acidic supernatant was subjected in succession to gel filtration on Sephadex G-50, ion-exchange chromatography on AG1-X2, and high voltage paper electrophoresis to yield small amounts of a tetrapeptide HA3-T2 (Table VII).

Of these eight peptides, two derived from AC-4 provided valuable overlaps. Twenty-seven cycles of Sequencer analysis of the 32-residue peptide AC4-TSS4 yielded a sequence corresponding to residues 601 to 627 (Table II) clearly overlapping CB15-CB8-CB2-CB24. Similar analysis of AC4-TS3 placed 10 residues (438-447) in sequence, clearly overlapping CB9-CB15.

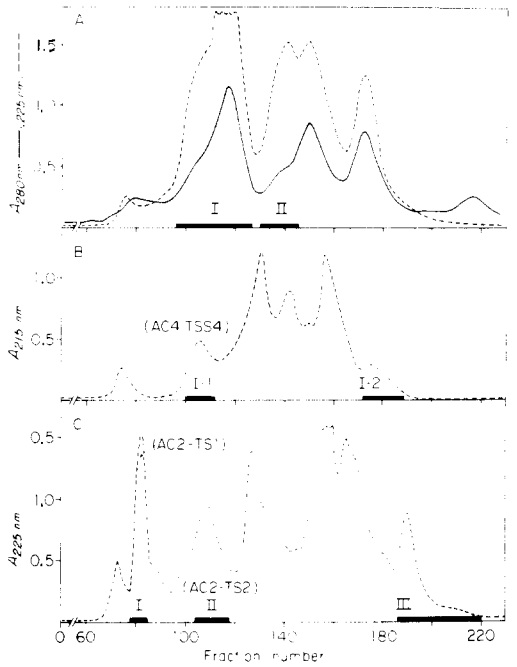

FIGURE 7: Separation of methionyl-containing peptides from tryptic digests of succinylated fragments (A) AC-4 (80 mg) and (C) AC-2 (60 mg). Panel B illustrates the separation of a staphylococcal protease digest of fraction I (26 mg) from A. Each separation utilized a column (2.5 × 200 cm) of Sephadex G-50 Superfine in 0.1 M NH_4HCO_3, pH 8.0. Fractions of 5 mL were collected at 30 mL/h and pooled as indicated.

The two peptides derived from AC-2 also contained several methionyl residues (five in AC2-TS1 and two in AC2-TS2) and a single arginyl residue. The peptide AC2-TS2 was treated with CNBr and found to contain CB1 and a fragment from the amino terminus of CB23 (residues 766-769), indicating the alignment CB1-CB23. This alignment was established by isolation of the carboxyl-terminal arginyl peptide (AC2-TS2-Th1) from a thermolytic digest of performic acid oxidized AC2-TS2 (Table VII). Its composition and partial sequence indicated the structure Leu-MetSO_2-His-His-Asp-Arg, clearly overlapping CB1-CB23 (Figure 3). The other tryptic peptide AC2-TS1 corresponded in composition to residues 649-713 (Table VII) and contained 5 methionyl residues. Extended Sequenator analysis of AC2-TS1 verified that this peptide began with valine-649, confirmed the carboxyl-terminal sequence of CB24 (through methionine-678), and continued with the sequence Lys-Phe-Met-Leu, thus placing CB3 after CB24. The last residue placed, leucine, must be derived from the amino terminus of either CB7 or CB1 since these are the only two fragments containing an amino-terminal leucine. However, it must belong to CB7 since CB1 is located in AC2-TS2 rather than in AC2-TS1, and in HA-3, rather than in HA-1. Thus, CB3 can only be linked to CB7. This was confirmed by fragmentation of AC2-TS1 with CNBr which generated pure CB3, CB7, and free arginine (in addition to an unresolved mixture which appeared to contain CB4, CB10, and part of CB24). Since there is only one arginyl residue in AC2-TS1, it must occur at the carboxyl terminus and the peptide must end with Met-Arg. To identify the source of this methionine, a pentapeptide (AC2-TS1-Th1) containing arginine was isolated from a thermolytic digest of performic acid-oxidized AC2-TS1 (Table VII). Sequence analysis of this pentapeptide (Table II) placed four residues, and the fifth was placed by composition in the sequence Ile-Phe-Gly-MetSO_2-Arg (residues 709-713). The first four residues overlap the carboxyl-terminus of CB10 and the arginine must represent the amino terminus of CB22, since it is the only CB fragment in AC-2 which contains an amino-terminal arginine.

A similar single residue overlap of CB10 and CB22 was suggested by the compositions of peptides derived by CNBr treatment of HA3-T1. These corresponded to CB7C, CB4, CB10, and again free arginine, presumably derived from the amino terminus of CB22.

Analysis of three other methionyl peptides (AC4-TSS1, AC4-TS2, and HA3-T2) yielded data consistent with the overlaps of CB17-CB18, CB18-CB9, and CB23-CB13, respectively (Table VII), but none of these peptides provides a convincing and independent proof of these alignments.

The strategy followed to prove the sequence of segment H_s is summarized in Figure 1. The complete sequence of phosphorylase is illustrated in Figure 3, which is arranged to focus on certain aspects of the proof presented herein, namely the sequence of fragment CB9 (residues 428-440), the sequence of the group of fragments comprising residues 604-841, and the four overlaps, CB17-CB18, CB18-CB9, CB9-CB15, and CB15-CB8.

Summary of Alignments of Cyanogen Bromide Fragments. The alignment of all 22 cyanogen bromide fragments in phosphorylase from the amino terminus to the carboxyl terminus is summarized as follows. Eight fragments have been placed in order in the amino-terminal 349 residues by Koide et al. (1978), ending with CB17 which contains the bond cleaved by subtilisin to generate segment H_s with CB17C at its amino terminus.

The next three fragments must be CB18, CB9, and CB15 since these are the only ones (besides CB17C) that are found in fragment HA-2. Of these, CB15 is carboxyl terminal because only its amino-terminal portion is found in HA-2. Fragment CB9 is linked to the amino terminus of CB15 by AC4-TS3 through the overlapping sequence Ile-Asn-Met-Ala-His-Leu. Although no satisfactory overlap of CB18 to CB9 was obtained, the sequence of the tryptic peptide AC4-TS2 (Met-Ser-Leu-Val-Glu-Glu-Gly-Ala-Val-Lys-Arg) is consistent with the cleavage of arginines 426 and 437. The only other Arg-Met sequence in phosphorylase which could yield an amino-terminal methionine following tryptic digestion of the succinylated protein occurs at residues 798-799 in CB23, but this fragment is in AC-2, not in AC-4. The linkage of CB17 to CB18 is supported by two observations: first, the composition of AC4-TSS1 (Leu, Met, Arg) is compatible only with a CB17-CB18 overlap (residues 348-350), and second, extended Sequencer analysis of AC-4 proceeded through fragment CB17Cac to reach aspartic acid-354. Thus, although sequences do not clearly overlap CB18 at either end, all data support its location between CB17 and CB9.

The next overlapping structure, derived from AC4-TSS4, links the unique carboxyl-terminal sequence of CB15 (Val-Met) to CB8-CB2-CB24, completing the alignment of 14 cyanogen bromide fragments in an uninterrupted chain of 678 residues.

The linkage of fragment CB24 to CB3 is established by analysis of AC2-TS1, in agreement with an overlapping peptide "T1P4" isolated by Forrey et al. (1971) from pyridoxyl-labeled phosphorylase. Peptide AC2-TS1 also provided a single residue overlap into CB7. This placement is supported by the fact that the only unaccounted fragment in HA-1 is CB7N and it must follow CB3.

The linkage CB7-CB4-CB10 is clearly demonstrated by the amino-terminal analysis of HA-3. The remaining linkages (CB10-CB22-CB1-CB23-CB13) are derived as follows. The sequence Leu-Met-His-His-Asp-Arg in AC2-TS2 establishes that CB1 precedes CB23 and the overlapping cyanogen bro-

mide peptide CB23-13 places CB1-CB23-CB13 in order. CB13 has been established as the carboxyl-terminal fragment. Thus, CB22, the only unplaced fragment in HA-3, must be between CB10 and CB1. Although no overlap between CB22 and CB1 was obtained, the contiguity of CB10 and CB22 is consistent with the sequence of AC2-TS1-Th1 (Ile-Phe-Gly-Met-Arg) since only CB22 has an amino-terminal arginine.

Discussion

This report describes the proof of the amino acid sequence of the carboxyl-terminal segment, H_s (residues 265-841), generated from rabbit muscle phosphorylase by subtilisin cleavage. The specific data include residues 428-440 (fragment CB9), residues 604-841, and their alignment (Figure 3) with residues 265-349 (Koide et al., 1978), residues 350-427 in CB18 (Hermann et al., 1978), and residues 441-603 in CB15 (Hermann et al., 1978).

As in the analysis of the amino-terminal 349 residues (Koide et al., 1978), some of the evidence presented is less certain than other. The alignment of the cyanogen bromide fragments CB17-CB18, CB3-CB7, and CB10-CB22 is based on marginal overlaps and no overlaps have been obtained for CB18-CB9, CB22-CB1, and CB23-CB13 (see Results). However, all the data are consistent with the proposed alignments and no other consistent arrangement of the 22 cyanogen bromide fragments has been found. The proposed alignments are fully consistent with the detailed crystallographic analyses of this molecule (see appendices). Weak points in our residue identifications are as follows: no corroborative analyses were performed in the placement of residues 436-437, 646-648, 748-752, 782-785, 792-794, and 810-832. Identifications of phenylthiohydantoin derivatives of asparagine-683, serines-746 and -750, aspartic acid-759, glutamic acid-778, arginine-794, and tryptophan-824 were considered to be tentative, but in each case residue placement was supported by the compositional analyses of a small peptide. Perhaps the weakest links in the sequence analysis relate to the identification of residues 749-759 in CB22, methionine-799 in CB23, and tryptophan-824 and arginine-832 in CB13. In fragment CB22 (Figure 6) the tryptic peptide T-6 (residues 753-758) is placed on the basis of compositional arguments and glutamine-753 is inferred from the indirect evidence for the presence of a cyclic pyrollidone. The placement of methionine-799 is deduced from the composition of CB23 and the observation that tryptic cleavage of this fragment at arginine-798 generates free homoserine. Tryptophan-824 and arginine-832 were observed only in very low yields during Sequencer analysis of CB13 (Table II). Placement of arginine-832 is consistent with the difference in composition between peptides T-8 and T-9 (Table III), though it is curious that the Arg-Leu bond in T-8 (residues 832-833) was not cleaved with greater efficiency.

Of the 18 cyanogen bromide fragments isolated by Saari & Fischer (1973), 15 were isolated from either segments H_s (the present communication) or L_s (Koide et al., 1978). Two additional fragments (CB14 and CB17) were recovered as recognizable subfragments following subtilisin treatment of the intact molecule. Fragment CB6 was not found herein and appears to have resulted in Saari & Fischer's experiments from anomalous cleavage by cyanogen bromide of CB20 as tryptophan-215 in the sequence Trp-Val-Asp-Thr-Gln-Val-Val-Leu-Ala-Met (assuming that only 70% of the Val-Val bond was hydrolyzed after 24 h in their experiments). Cleavage by brominating agents at tryptophan has been reported for other proteins (Savige & Fontana, 1977). Thus the previous generation of fragment CB6 may have resulted from contaminants such as free bromine which were absent in our work.

Nine small cysteinyl peptides were isolated from phosphorylase by Zarkadas et al. (1968). The sequences of five of these are identical with those obtained here, namely, those containing cysteines 108, 317, 371, 44, and 782. Two others correspond to peptides containing cysteines 494 and 579, except that residue 495 is asparagine rather than aspartic acid and residue 575 is glutamine rather than glutamic acid. An eighth peptide was originally assigned the sequence Asn-Gln-Lys-Cys-Gly-Gly, later referred to as Asn-Glu-Lys-Ile-Cys-Gly-Gly (peptide B_1, Avramovic-Zikic et al., 1970), and must correspond to our sequence Asn-Gln-Lys-Ile-Cys-Gly-Gly surrounding cysteine-171. The ninth cysteine residue was reported by Zarkadas et al. (1968) to occur in the sequence Ala-Cys-Ala, which was later extended to Ala-Cys-Ala-Phe (peptide N of Battell et al., 1968); we find cysteine-142 within the sequence Ala-Ala-Cys-Phe. Allowing for these discrepancies, all of the cysteinyl peptides of Zarkadas et al. (1968) can be placed in our sequence. The relationship of four of these residues to the catalytic function and aggregation behavior of phosphorylase has been reviewed by Koide et al. (1978).

In another region of the molecule, our conclusions are at variance with a 42-residue sequence reported by Forrey et al. (1971) which includes the lysyl residue to which the cofactor pyridoxal 5′-phosphate had been covalently attached by $NaBH_4$ reduction. Their amino-terminal sequence was based on a 33-residue chymotryptic peptide which corresponds to residues 648-680 (overlapping fragments CB24 and CB3), except that we find glutamic acid and asparagine rather than glutamine and aspartic acid at residues 671 and 677. These data identify lysine-679 as the site of pyridoxal phosphate attachment. However, Forrey et al. (1971) also isolated from a cyanogen bromide digest a pyridoxyl-labeled peptide reported to contain an uncleaved homoseryl residue and having a composition equal to the sum of those of CB3 and CB5. They suggested, therefore, that peptide CB5 would follow CB3. However, the data of Koide et al. place CB5 in segment L_s (residues 92-99) between CB14 and CB12 (by analysis of fragments AC-1 and L_s-T-1), whereas CB3 is located in segment H_s (residues 679-681) between CB24 and CB7 (by analysis of peptide AC2-TS1). Furthermore, peptide CB3 is found in fragments HA-1 and AC-2 (Table V), whereas CB5 is not. It is difficult to explain these earlier data, especially since in our sequence CB5 is preceded by the sequence Gln-Phe-Tyr-Met (in CB14) which does *not* resemble CB3 (Lys-Phe-Met). A recent analysis of the pyridoxal binding site in potato phosphorylase reveals a homologous relationship with the rabbit enzyme, which is in accordance with our sequence in this region of the molecule (Nakano et al., 1978).

Since the crystallization of rabbit muscle phosphorylase by Green & Cori (1943), 35 years of kinetic, chemical and physical investigations have provided a wealth of information on the nature of this enzyme. The determination of the complete primary structure of this 97 412-dalton molecule, in conjunction with the electron density maps obtained in Edmonton for phosphorylase *a* (Sygusch et al., 1977) and in Oxford for phosphorylase *b* (Johnson et al., 1977; Weber et al., 1978) now permits the elaboration of a detailed model to explain the structure-function relationship of the enzyme. Kinetic studies have revealed the presence of a multiplicity of regulatory sites subject to covalent modifications as well as complex modulation by metabolites or allosteric effectors. Covalent control is subject to the opposing influences of phosphorylase kinase and phosphatase which are responsible

for the reversible interconversion of inactive phosphorylase *b* and active phosphorylase *a*. However, the phosphorylation and dephosphorylation of the protein are not solely determined by these two regulatory enzymes, but by a number of enzymes acting successively on one another. Furthermore, this cascade of enzymatic reactions is closely linked to other physiological processes such as hormone stimulation and the nerve impulse that triggers contraction. Allosteric regulation is exercised by a number of positive (e.g., AMP, IMP, glucose 1-phosphate) and negative (e.g., ATP, adenosine, caffeine, glucose, glucose 6-phosphate, UDPG, etc.) effectors (for reviews, see Fischer et al., 1971; Graves & Wang, 1972).

Both the nature of the active site of the enzyme and its location have been matters of particular interest and controversy. The reaction catalyzed by phosphorylase proceeds with complete retention of configuration which might suggest a double displacement mechanism involving a covalent intermediate with the enzyme. Such a glucosyl intermediate is produced in the sucrose phosphorylase reaction which also proceeds with retention of configuration; it has been isolated and displayed the β configuration expected for a double displacement mechanism (Voet & Abeles, 1970). Also, the predicted isotope exchange occurs in the absence of fructose. No such glycosyl intermediate has ever been found in glycogen phosphorylase and no half-reaction or isotopic exchange has been observed with the muscle enzyme.[2]

All glycogen phosphorylases so far investigated contain stoichiometric amounts of pyridoxal phosphate which has therefore been suspected as a participant in the catalytic mechanism. However, while removal of the cofactor from the enzyme leads to total inactivation, reduction with sodium borohydride, which fixes the cofactor to the protein, results in a product which is still enzymatically active. All other classical B_6-enzymes in which pyridoxal phosphate is directly involved in catalysis are totally inactivated by this kind of treatment. It was therefore assumed that, if pyridoxal phosphate is indeed a direct participant, it would have to function in a way different from all other pyridoxal phosphate containing enzymes (Fischer et al., 1970, 1971). Positioning of the cofactor at lysine-679 in the sequence (Titani et al., 1977) has facilitated the location of the cofactor in the three dimensional structure by providing a basis on which the polypeptide chain could be traced from cysteine-782 to the ϵ-pyridoxallysine-679 in both phosphorylase *a* (Sygusch et al., 1977) and in phosphorylase *b* (Weber et al., 1978).

Although the data are still indirect, there now appears to be overwhelming evidence that pyridoxal phosphate is directly involved in catalysis. First, as indicated above, the pyridoxal phosphate binding site in muscle, yeast (Lerch & Fischer, 1975), and potato phosphorylase (Nakano et al., 1978) has been remarkably conserved. By contrast, the phosphorylated site involved in the covalent control of yeast phosphorylase is totally different from that observed in the rabbit (Lerch & Fischer, 1975); it is entirely absent in the higher plant enzyme. It is almost inconceivable that the pyridoxal phosphate binding site would have been conserved during the approximately 1.5 billion years of divergent evolution of *Saccharomyces* and mammals if it were not crucial to the nature of the enzyme, that is, if it were not directly involved in catalysis. Second, phosphorylase binds glucose (a competitive inhibitor), inorganic phosphate and glucose 1-phosphate (two of its substrates) in close proximity to the pyridoxal phosphate binding site in phosphorylase *a* (Sygusch et al., 1977); the same observation was made for glucose 1-phosphate in phosphorylase *b* (Johnson et al., 1977; Weber et al. 1978), and both crystallography groups have now suggested that this particular site must represent the active center of the enzyme. Previously, a distant site located in the area of tyrosine-155, which binds glucose 1-phosphate with an even greater affinity, seemed to be a more likely candidate, but it now appears to represent a ligand binding site involved in allosteric control. Third, the NMR studies of Feldmann & Helmreich (1976) and Feldman & Hull (1977) have quite convincingly demonstrated the involvement of the phosphate group of pyridoxal phosphate in all transitions from inactive to active muscle phosphorylase so far observed. These data are compatible with a proton donor-acceptor function of this phosphate group during catalysis.

The X-ray crystallographic studies also indicate that the contact region between the two protomers is in the amino-terminal region of the molecule; the chemical studies of Battell et al. (1968) had shown that cysteine residues 108 and 142 become exposed in the course of disaggregation to the protomers. Crystallographic analyses also revealed that oligosaccharide analogues of glycogen (maltoheptaose and maltotriose) both bind strongly to a site 25–50 Å from the active center and more weakly to the active site itself. The precise location of the strong binding site has not yet been identified in the primary structure. Fletterick et al. (1976) have suggested that it serves to immobilize phosphorylase on the glycogen particle, while allowing ready access of high local concentration of oligosaccharides to the active site and of the various ligands to their respective regulatory sites. In the domain implicated in the major allosteric control of the enzyme, there is a good correlation between the sequence data and the difference Fourier maps. Both point to tyrosine-155 in the region involved in AMP binding. It should be noted, however, that this specific residue has been tagged by using the very large affinity label 8-[*m*-(*m*-fluorosulfonylbenzamido)benzylthio]adenine (Anderson & Graves, 1973; Anderson et al., 1973); by itself it might have no physiological role. That is, it is to be expected that another residue would have been covalently modified by a different analogue of AMP. Indeed, chemical modification studies have implicated an arginyl residue in the AMP binding site (Li et al., 1977), but is is not yet clear which specific one is involved. Further correlation between the primary structure and electron density maps of this complex enzyme should reveal in greater detail the molecular features of the active center as well as those of the different ligand binding sites in the various regulatory domains.

Most of the regulatory sites appear to be located in the amino-terminal segment (L_S) of the molecule, whereas the catalytic site is buried in the heavy segment (H_S). These two domains are loosely linked by a segment readily accessible to proteolytic enzymes. It is therefore conceivable that at one time the molecule existed as a two-subunit (regulatory-catalytic) heteropolymer as found for many regulatory enzymes including the cAMP-dependent protein kinase (Krebs, 1972) and phosphorylase kinase (Cohen, 1974). Perhaps by gene translocation the two subunits became covalently linked, accounting for the unusually large size of the enzyme subunit. Alternatively, of course, the enzyme could have existed originally as a single, large catalytic molecule without a regulatory domain—as observed, for instance, with potato phosphorylase (Lee, 1960). A regulatory site would then have been created later in the course of evolution, in response to physiological demands.

[2] Loss of ^{18}O from the phosphoryl oxygen bridge of glucose 1-phosphate at a rate corresponding to that of starch synthesis has been observed for potato phosphorylase in the presence of cyclodextrin which cannot serve as a real substrate because it lacks a free hydroxyl group in position 4 (Kokesh & Kakuda, 1977).

Acknowledgments

The authors are grateful to Rosa Mae Macdonald for her assistance with the high performance liquid chromatography and to Richard R. Granberg for the amino acid analyses. It is also a pleasure to acknowledge informative exchanges of information with Drs. R. Fletterick and N. Madsen (University of Alberta) and with Dr. L. H. Johnson (University of Oxford) throughout this work (see appendices).

References

Anderson, R. A., & Graves, D. J. (1973) *Biochemistry 12,* 1895.

Anderson, R. A., Parrish, R. F., & Graves, D. J. (1973) *Biochemistry 12,* 1901.

Avramovic-Zikic, O., Smillie, L. B., & Madsen, N. B. (1970) *J. Biol. Chem. 245,* 1558.

Battell, M. L., Zarkadas, C. G., Smillie, L. B., & Madsen, N. B. (1968) *J. Biol. Chem. 243,* 6202.

Bennett, J. C. (1967) *Methods Enzymol. 11,* 330.

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

Cohen, P. (1974) *Biochem. Soc. Symp. 39,* 51.

Cohen, P., Saari, J. C., & Fischer, E. H. (1973) *Biochemistry 12,* 5233.

Feldmann, K., & Helmreich, E. J. M. (1976) *Biochemistry 15,* 2394.

Feldmann, K., & Hull, W. E. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74,* 856.

Fischer, E. H., Pocker, A., & Saari, J. C. (1970) in *Essays of Biochemistry* (Campbell, P. N., & Dickens, F., Eds.), p 23, Academic Press, New York, N.Y.

Fischer, E. H., Heilmeyer, L. M. G., & Haschke, R. H. (1971) *Curr. Top. Cell. Regul. 4,* 211.

Fletterick, R. J., Sygusch, J., Semple, M., & Madsen, N. B. (1976) *J. Biol. Chem. 251,* 6142.

Forrey, A. W., Sevilla, C. L., Saari, J. C., & Fischer, E. H. (1971) *Biochemistry 10,* 3132.

Graves, D. J., & Wang, J. H. (1972) *Enzymes, 3rd Ed. 7,* 435.

Green, A. A., & Cori, G. T. (1943) *J. Biol. Chem. 143,* 21.

Hermann, J., Titani, K., Ericsson, L. H., Wade, R. D., Neurath, H., & Walsh, K. A. (1978) *Biochemistry 17* (second of three papers in a series in this issue).

Hirs, C. H. W. (1967) *Methods Enzymol. 11,* 197.

Horn, M. J., & Laursen, R. A. (1973) *FEBS Lett. 36,* 285.

Johnson, L. N., Weber, I. T., Wild, D. L., Wilson, K. S., & Yeates, D. G. R. (1977) *Fed. Eur. Biochem. Soc., Meet., 11th,* 185.

Koide, A., Titani, K., Ericsson, L. H., Kumar, S., Neurath, H., & Walsh, K. A. (1978) *Biochemistry 17* (first of three papers in a series in this issue).

Kokesh, F. C., & Kakuda, Y. (1977) *Biochemistry 16,* 2467.

Konigsberg, W., & Hill, R. J. (1962) *J. Biol. Chem. 237,* 2547.

Krebs, E. G. (1972) *Curr. Top. Cell. Regul. 5,* 99.

Lee, Y. P. (1960) *Biochim. Biophys. Acta 43,* 18.

Lerch, K., & Fischer, E. H. (1975) *Biochemistry 14,* 2009.

Li, E. C. Y., Fletterick, R. J., Sygusch, J., & Madsen, N. B. (1977) *Can. J. Biochem. 55,* 465.

Nakano, K., Wakabayashi, S., Hase, T., Matsubara, H., & Fukui, T. (1978) *J. Biochem. Jpn. 83,* 1085.

Ryle, A. P., Sanger, F., Smith, L. F., & Kitai, R. (1955) *Biochem. J. 60,* 541.

Saari, J. C., & Fischer, E. H. (1973) *Biochemistry 12,* 5225.

Savige, W. E., & Fontana, A. (1977) *Methods Enzymol. 47,* 459.

Smith, I. (1960) in *Chromatographic and Electrophoretic Techniques* (Smith, I., Ed.), Vol. 1, p 82, Interscience, New York, N.Y.

Sygusch, J., Madsen, N. B., Kavinsky, P. J., & Fletterick, R. J. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74,* 4757.

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

Titani, K., Ericsson, L. H., Neurath, H., & Walsh, K. A. (1975) *Biochemistry 14,* 1358.

Titani, K., Koide, A., Hermann, J., Ericsson, L. H., Kumar, S., Wade, R. D., Walsh, K. A., Neurath, H., & Fischer, E. H. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74,* 4762.

Voet, J. G., & Abeles, R. H. (1970) *J. Biol. Chem. 245,* 1020.

Weber, I. T., Johnson, L. N., Wilson, K. S., Yeates, D. G. R., Wild, D. L., & Jenkins, J. A. (1978) *Nature (London) 274,* 433.

Wikler, M., Titani, K., & Putnam, F. W. (1970) *J. Biol. Chem. 245,* 2158.

Yamada, S., & Itano, H. A. (1966) *Biochim. Biophys. Acta 130,* 538.

Zarkadas, C. B., Smillie, L. B., & Madsen, N. B. (1968) *J. Mol. Biol. 38,* 245.

Appendix I: Crystallographic Analysis of Phosphorylase *a* at 2.5 Å Resolution, a Comment on the Chemical Sequence

Since our last communication of the crystallographic analysis of phosphorylase *a* at 3.0 Å resolution (Fletterick et al., 1976), an additional "heavy-atom" derivative has been used and the phasing extended to 2.5-Å resolution. These improvements of the electron density map have enabled us to use the chemical sequence to solve the structure and produce a reliable set of atomic coordinates for the enzyme. This work will be published in detail later. In this report we discuss the crystallographic results as they apply to confirmation of the chemical sequence about the 21 methionines and therefore the overlapping sequences used to align the 22 cyanogen bromide peptides.

The electron density map used to measure the atomic coordinates is of good quality. The phase calculation for about 25 000 reflections to 2.5-Å resolution gave a figure-of-merit (Sygusch, 1977) of 0.70. The map was stacked on Plexiglass in sections at intervals of 0.8 Å along the Z direction using a scale of 0.7 cm/Å. The map was interpreted by "fitting" the electron density while visually guided with Lab-Quip models (1.0 cm/Å). The atomic positions were marked in the map sections and recorded graphically. The approximate coordinates were justified to conform to expected values of molecular geometry using the method of weighted restraints (Hermans & McQueen, 1974). The model was then modified with a computer program which adjusted the coordinates so as to increase the electron density at the atomic positions. Three cycles of the above two operations produced stable coordinates of acceptable geometry which were reasonably well placed in the electron density.

Our detailed examination of this map completely confirms the ordering of the cyanogen bromide fragments with only some instances of weak corroboration. These are due to regions in the electron density map corresponding to the surface of the molecule where the polypeptide chain loops into the solvent and no connected electron density is observed. This situation applies to the following peptide segments: 1–4; 17–21; 257–261; 315–324.

Aside from these regions there are no other breaks in main chain density with the exception of the three amino acids, 384,

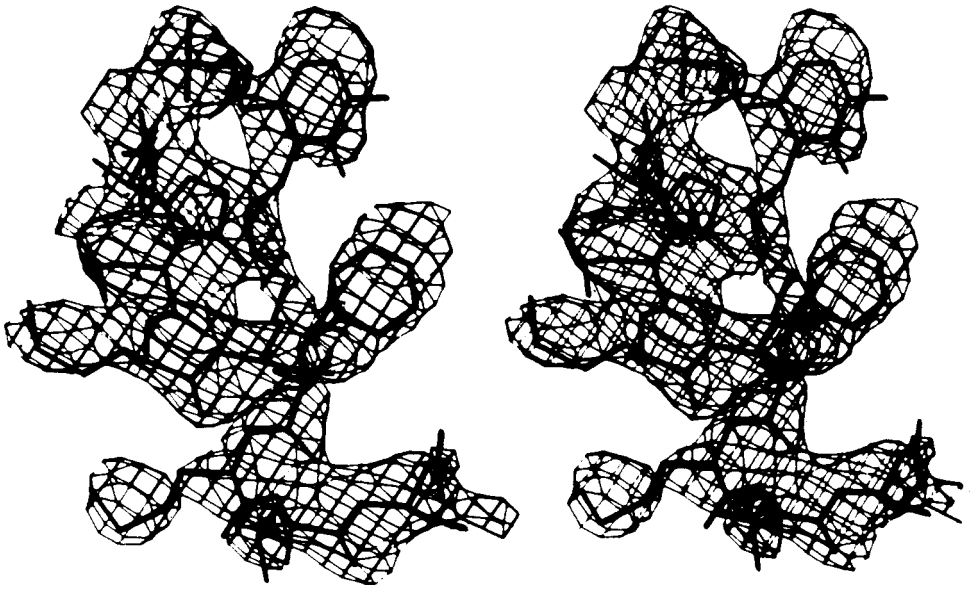

FIGURE A-1: A stereo view of the chain segment from 675 to 683 (TGNMK(Plp)GMLN) showing the atomic positions within the experimental electron density map. The photograph was made from a MMS-X graphics computer. Plp denotes pyridoxal 5′-phosphate covalently attached to lysine-679.

TABLE A-I: Evaluation of the Native Map at 17 Subsequences Containing Methionine.

sequence start	sequence	av electron density[a] main chain	av electron density[a] side chain	quality of chain recognition
86	SLQFYMGRTLQ	39	29	good
94	TLQNTMVNLAL	47	38	excellent
114	QLGLDMEELEE	35	27	poor[b]
142	CFLDSMATLGL	41	34	good
171	CGGWQMEEADDW	37	36	good
219	QVVLAMPYDTP	47	32	good
236	NVVNTMRLWSA	47	36	good
344	AIPELMRVLVD	53	44	excellent
422	DRLRRMSLVEE	54	30	fair
435	VKRINMAHLCI	51	36	good
598	VPRTVMIGGKA	44	29	good
609	APGYHMAKMIIKLI	59	30	good
673	SGTGNM(Plp)FML-NGAL	54	35	excellent
686	LTIGTMDGANVE-MAEEA	52	42	excellent
707	FFIFGMRVEDVD	47	32	good
758	KDIVNMLMHHDRF	53	31	good
794	REWTRMVIRNI	48	29	fair

[a] Average electron density over all main chain and side chain atoms in each subsequence is expressed in arbitrary units. The maximum background electron density of the map is 10. [b] Lead acetate binds in this region and is one of the heavy-atom derivatives used to calculate the multiple isomorphous replacement phases. The lead interaction causes substantial chain movement at the binding site; consequently the electron density at 120–124 is obscured in the native map.

546, and 564. Table A-I contains a summary of the quality of the map at those regions corresponding to 21 short subsequences containing methionine residues. The level of confidence with which we are able to corroborate the chemical sequence is not only based on the shape and magnitude of the side chain electron density, but also the knowledge of the tertiary structure and thus the chemical environment at the residues in question. The quality of the map is such that in most regions about 40% of the amino acid side chains can be identified correctly without the sequence. Within the region from 700 to 810, about 70% of the side chains can be "sequenced" directly from the map. The completion of the correct chain tracing and measurement of the atomic coordinates were, however, entirely dependent on the amino acid sequence. The chain trace in the C-terminal half is slightly modified from that published earlier (Fletterick et al., 1976).

Figure A-1 shows a stereo view of the structure of the peptide chain from 675–683 and the corresponding electron density. This sequence includes the pyridoxal 5′-phosphate covalently bound to the ϵ-nitrogen of lysine-679. These coordinates are unrefined except for the technique described above. The agreement among the model, the sequence, and the observed electron density map is seen to be satisfactory. We conclude that the ordering of cyanogen bromide fragments is correct. Further, there are very few (less than ten) cases of individual amino acids for which the X-ray results conflict with the chemical sequence. These conflicts are ascribed to probable error in the electron density map.

References

Fletterick, R. J., Sygusch, J., Semple, H., & Madsen, N. B. (1976) *J. Biol. Chem. 251,* 6142–6146.

Hermans, J., & McQueen, J. (1974) *Acta Crystallogr., Sect. A 30,* 730–739.

Sygusch, J. (1977) *Acta Crystallogr., Sect. A 33,* 512–518.

Appendix II: Assignment of the Amino Acid Sequence to the Crystal Structure of Glycogen Phosphorylase *b*

The crystal structure of glycogen phosphorylase *b* has been solved at 3-Å resolution using phases determined from two heavy atom derivatives (Johnson et al., 1978). From an examination of a small scale map (0.5 cm ≡ 1 Å), the amino acid sequence (Titani et al., 1977) has been fitted to the electron density for residues 20–311, 349–449, and 643–810. While these assignments have still to be verified by molecular model building, the ease with which key residues, such as those containing aromatic groups or sulfur atoms, could be identified provides confidence in our interpretation. Of the remainder of the chain, the following comments can be made. Residues 1–19 appear flexible in phosphorylase *b* (Weber et al., 1978) and cannot be located. There is a break in the continuity of the

electron density of about 4–5 residues after Lys-311. The fit of the sequence 317 to 348 appears plausible but it has not yet been assigned in detail. Likewise the region between 450 and 642 is in continuous density with the right number of residues but assignments have not yet been made. The interpretation of the course of the chain from residue 810 onwards differs markedly from that for phosphorylase *a* and assignments have still to be made in detail. The loop between residues 159–183 is in weak density, possibly due to some disorder caused by our platinum heavy atom derivative.

The assignments to date allow us to confirm the alignments of the following cyanogen bromide fragments: CB14–CB5 (Met-91), CB5–CB12 (Met-99), CB12–CB16 (Met-119), CB16–CB21 (Met-147), CB20–CB11 (Met-224), CB11–CB17 (Met-241), CB18–CB9 (Met-427), CB9–CB15 (Met-440), CB24–CB3 (Met-678), CB3–CB7 (Met-681), CB7–CB4 (Met-691), CB4–CB10 (Met-698), CB10–CB22 (Met-712), CB1–CB23 (Met-765), CB23–CB13 (Met-799). The overlap of CB21–CB20 (Met-176) is in weak density but since large regions of both fragments have been assigned there seems little reason to doubt the sequence. The overlap of CB17–CB18 is in an ambiguous region and it is hard to identify the C terminus of CB17. However the number of residues in this region is approximately in concordance with the sequence information. The density for Met-763 is weak for a methionine side chain but the alignment CB22–CB1 appears plausible in terms of the number of residues.

There are two regions where the electron density map appears to require fewer residues than those given in the sequence. The first is the region 360–364 (Trp-Asp-Lys-Ala-Trp). The two tryptophans are clearly visible and Trp-364 makes contact to glucose 1-phosphate bound at the active site (Weber et al., 1978). However, the density between the two tryptophans appears to indicate 2 rather than 3 residues. Secondly in the region 776–779 (Tyr-Glu-Glu-Tyr), the density suggests one rather than two glutamic acid residues.

In summary, at the present preliminary stage of our analysis, our results are consistent with the sequence data. In particular, binding studies with ATP have shown that Tyr-155 which can be covalently modified by an analogue by AMP (Anderson et al., 1973) is in the right position to form this link. At Lys-679 there is good density for the pyridoxal phosphate and this site is close (8 Å) to the site at which glucose 1-phosphate binds tightly (Weber et al., 1978). The two major mercury binding sites for the heavy atom derivative ethylmercury thiosalicylate are at Cys-782 and Cys-108, while the minor site which is also close to the major platinum site is at Cys-171.

References

Anderson, R. A., Parrish, R. F., & Graves, D. J. (1973) *Biochemistry 12,* 1901–1906.

Johnson, L. N., Weber, I. T., Wild, D. L., Wilson, K. S., & Yeates, D. G. R. (1978) *Fed. Eur. Biochem. Soc., Meet., 11th,* 185–194.

Titani, K., Koide, A., Hermann, J., Ericsson, L. H., Kumar, S., Wade, R. D., Walsh, K. A., Neurath, H., & Fischer, E. H. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74,* 4762–4766.

Weber, I. T., Johnson, L. N., Wilson, K. S., Yeates, D. G. R., Wild, D. L., & Jenkins, J. A. (1978) *Nature (London) 274,* 433.

Transcription of Histone-Covered T7 DNA by *Escherichia coli* RNA Polymerase†

P. Williamson*,‡ and G. Felsenfeld

ABSTRACT: Purified core histones (H2A, H2B, H3, and H4) and bacteriophage T7 DNA have been reconstituted to form a nucleoprotein complex, and the properties of this complex as a template for transcription by *Escherichia coli* RNA polymerase have been studied. At low ionic strength, RNA chain elongation rates are slow, and the chains produced even after long incubation are short. At higher salt concentrations, chain-elongation rates approach those on naked DNA. Since the salt concentrations used are not in themselves sufficient to dissociate the histones from the DNA, some mechanism must exist that permits passage of the polymerase through histone-covered regions.

No understanding of transcription as it occurs in eukaryotes can be complete without some idea of the role played by the structure of the chromatin template in the various processes which finally give rise to specific RNA products. Thus, studies of the transcription of chromatin must eventually give a description of the effects of the nucleosomes on RNA chain initiation, propagation, termination, and processing. That these effects may be of some importance is indicated by evidence suggesting that transcriptionally active gene sequences are associated with some form (perhaps modified) of nucleosome (Foe et al., 1976; Lacy and Axel, 1975; Weintraub and Groudine, 1976; Camerini-Otero et al., 1978). In order to approach this problem, we have made use of a model system in which the effects of nucleosomes on chain elongation in transcription can be examined systematically. In this system, we have reconstituted the purified histones of the nucleosome core onto bacteriophage T7 DNA and have used conditions under which initiation and propagation by *E. coli* RNA polymerase can be measured separately. We find that both processes are affected. At low ionic strength, propagation is slowed, and eventually halted, by nucleosomes; these inhibitory

† From the Laboratory of Molecular Biology, National Institute of Arthritis, Metabolism, and Digestive Diseases, National Institutes of Health, Bethesda, Maryland 20014. *Received May 16, 1978; revised manuscript received September 7, 1978.* A preliminary account of this work has been published (Camerini-Otero et al., 1978).

‡ Present address: Department of Biology, Amherst College, Amherst, Mass. 01002.